THE UNIVERSITY *of* EDINBURGH

Genetic control of anthocyanin pigmentation in *Antirrhinum* flowers

Tanyarat Khongkhuntian

A thesis submitted for the degree of

Doctor of Philosophy

The University of Edinburgh

2012

Abstract

The genus *Antirrhinum* (commonly known as snapdragons) contains more than twenty-five recognised species. The genus has been divided into three morphological subsections: *Antirrhinum*, *Streptosepalum* and *Kickxiella* (Rothmaler, 1956). One of the major characteristics distinguishing the three subsections is flower colour. Most species in subsection *Antirrhinum* have dark pink or yellow flowers, *Kickxiella* species are white or pale pink and *Streptosepalum* species have yellow or pale pink flowers. All *Antirrhinum* species can be crossed to produce fertile hybrids which allow the genes that underlie their differences to be identified.

I used quantitative trait locus (QTL) analysis on hybrids of *A. majus* (dark magenta flowers) and *A. charidemi* (pale-pink flowers) to map genomic regions underlying differences in flower colour. This identified two major-effect loci, in Linkage Group 3 (LG3) and LG7, that explained most of the differences between these species. I used near-isogenic lines (NILs) to further test involvement of two candidate genes - *Rosea* (*Ros*) in LG3, which encodes a regulator of the anthocyanin biosynthesis pathway (ABP) and *Incolorata* (*Inc*) in LG7 which encodes a rate-limiting enzyme of the ABP. In both cases, the *A. majus* allele increased pigmentation. Sequence differences between *Ros* alleles of *A. majus*, *A. charidemi* and *A. molle* (a *Kickxiella* species with white flowers) suggest that *A. molle* carries a *ros* loss-of-function mutation and that a transposon insertion in the *ROS* promoter might contribute to differences in expression between *A. majus* and *A. charidemi*. *Ros* genotypes were found to be strongly correlated with pigmentation in the corolla tube in *A. majus* x *A. charidemi* hybrids, and to a lesser extent with corolla lobe pigmentation, although NILs suggested that *ROS* did not correspond to the major-effect QTL indentified in LG3. I also mapped a minor-effect QTL for tube pigmentation to a region of LG4 containing the ABP structural gene *Candica*. Analysis of NILs revealed that *Inc* was not the second major-effect QTL mapped to LG7, although sequence differences were detected between *Inc* alleles of *A. majus* and *A. charidemi*. I was further able to narrow down the region containing the second LG7 major-effect QTL to an interval of 11 cM, between two molecular markers, which could be used to determine the likely QTL genotypes of segregating NILs. Surprisingly, several ABP genes, particularly *Nivea*, *Inc* and *Pallida*, were expressed at higher levels in pale flowers that were homozygous for the *A. chardemi* QTL allele than in their dark flowered siblings that carried an *A. majus* allele. This suggests that ABP genes might be up-regulated in pale flowers as part of a negative feedback mechanism. Two potential roles of the LG7 QTL are considered 1) its requirement for anthocyanin modification or transport to the vacuole, so that a build-up of cytosolic anthocyanins or their break-down products in pale flowers increases structural gene expression but cannot compensate for the overall reduction in anthocyanin, or 2) a role in promoting production of flavonols at the expense of anthocyanins.

Acknowledgments

This work would not have been done without many people's support and help. Most of all, my kind supervisor, Andrew Hudson, who devoted himself to help me through my PhD, in both the laboratory and the writing stage. Without his guidance, my project would not have been successful. I am also grateful to my second supervisor, Catherine Kidner, for her comments and suggestions.

I would like to thank Cathie Martin, Kathy Schwinn who kindly gave me the unpublished DNA sequence of *Mut* and *WD40* , and suggestion for the expression results, Francesca Quattracchio for her advice with the expression results and Nicholas Langlade for his picture of the F_2 *A.majus x A. charidemi* population. Xingzhong Feng for all his markers data of *Antirrhinum* linkage and DNA samples. Yong-Biao Xue gave me the data sequence of *A. majus* genome to work on.

I am also grateful to other PIs, Justin Goodrich, Gwyneth Ingram, Stephen Fry and Peter Doener for advice and suggestions in many stages of this project.

I am also grateful to ex-members of the Hudson group, who had shared the unpublished data, DNA samples, plant samples, and gave me useful advice. Manuela Costa shared her unpublished expression results of *Rosea.* Suxin Yang shared *ML-1* marker data. Chia Ching Yang shared his linkage map data in *A .majus* x *A. molle.*

Special thanks for Joanna Critchley for sharing her unpublished discovery for *Rosea* 5'UTR and all the advice of *Antirrhinum* data collecting, propagating method and her big help in the genotyping and AFLP stage. Her encouragement and support helped me through the tough stages of my project and my life.

I would like to thank Andrew Waters who prepared the *E. coli* competent cells for me and Janice Miller for her advice on TLC.

I have to say thanks to all previous and current members of the Hudson group, the Goodrich group and the big lab for support and help. Special thanks to Hayley McCulloch and Byung-Ho Hwang for their encouragement during my thesis writing. I would like to say thank you to Yvette Wilson, Poay Lim, Abigail Harter, Zafar Malik, Erica de Leau and other colleagues in the 'big lab' for their help in lab and friendship.

My thanks to all Thai student friends, especially to Kelwalin Dhanasarnsombut; her encouragement helped me through the final stage of thesis writing.

I would like to say thank you for Arshai Moghtaderpour for his final stage proof reading to my entire thesis and for all of his support during my study.

My study at the University of Edinburgh was funded by Thai government Science and Technology Scholarship.

Last, I am grateful to my family for their love, support and encouragement for my study, especially my beloved mother, who I know would be happy watching over me now.

Declaration

I hereby declare that I am the author of this thesis. The work presented is my own original work and I have not previously submitted this work in its entirety, or in part, for any other degree or professional qualification

…….. ……..

Tanyarat Khongkhuntian Date

Table of contents

List of figures

List of tables

Abbreviations

5'UTR	5' Untranslated Region
ABP	Anthocyanin Biosynthesis Pathway
AFLP	Amplified Fragment Length Polymorphism
bHLH	Basic Helix-Loop-Helix protein
bp	Base Pairs
BSA	Bovine Serum Albumin
CAPS	Cleaved Amplified Polymorphisms
cM	Centimorgan
CTAB	Cetyltrimethyl Ammonium Bromide
DTT	Dithiothreitol
EDTA	Ethylenediaminetetracetic Acid
HCl	Hydrochloric Acid
iPCR	Inverse Polymerase Chain Reaction
indel	Insertion-deletion Polymorphism
kb	Kilobase
PCR	Polymerase Chain Reaction
pers.comm.	Personal Communication
MBW	MYB-bHLH-WD complex
NIL	Near Isogenic Line
QTL	Quantitative Trait Locus
SNP	Single Nucleotide Polymorphism
TBE	Tris/Borate/EDTA Buffer
Tris	Tris(hydroxymethyl)aminomethane

Chapter 1
Introduction

1.1 *Antirrhinum*: A model genus

Antirrhinum has been widely known as a flowering ornamental plant for several centuries. Southern Europe contains more than twenty recognised species and subspecies of *Antirrhinum* (commonly known as snapdragons) including the cultivated *A. majus*. According to their morphological characteristics and habitats, they can be divided into three subsections: *Antirrhinum*, *Streptosepalum* and *Kickxiella* (Rothmaler, 1956). The subsection *Antirrhinum* includes the close relatives of *A. majus* and consists of species with similar upright growth, large organs, and pink or yellow flowers. This group is geographically widespread and grows in a variety of habitats, mostly found in north-eastern Spain to south-western France. Members of subsection *Kickxiella*, in contrast, are usually restricted to rock faces and walls. They are also smaller, typically prostrate in habit, and have small white or pale pink flowers. They are geographically more restricted, and many are endemic to particular mountain regions. The two members of subsection *Streptosepalum* are pale- or yellow-flowered, of upright habit, and grow in hedges and rocky outcrops in northern Spain and Portugal (Schwarz-Sommer *et al.,* 2003) (Figure 1b).

Antirrhinum is a member of the asterid clade of flowering plant. Within the clade, *Antirrhinum* belongs to the order Lamiales and has recently been placed in the family Plantaginaceae (sometimes known as Veronicaceae) by the revision of the classical family Scrophulariaceae using DNA sequence variation (Olmstead *et al*., 2001). Normally the generic epithet "*Antirrhinum*" is used to refer to the monophyletic group of Old World (European) perennials with the diploid chromosomes number of 2n = 16. However, it is sometimes extended to include sister groups with different chromosome numbers, geography and habits, such as the New World *Sairocarpus* (2n = 32), *Howeliella* (2n = 32), *Neogaerrhinum* (2n = 30) and the Old World annual

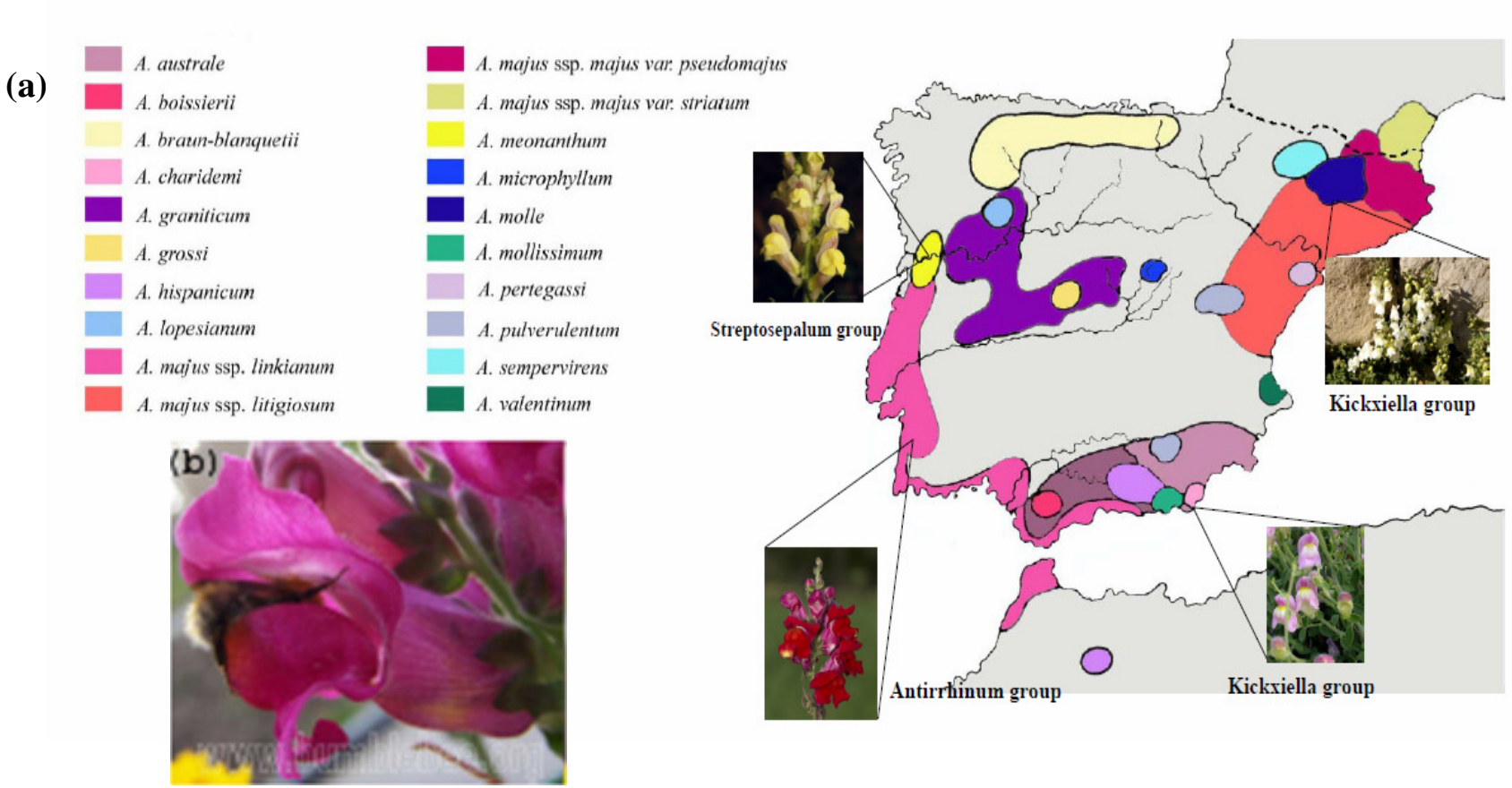


Figure 1. *Antirrhinum* distribution in southern Europe and North Africa and its primary pollinators. (a) The distribution of *Antirrhinum* species across the Iberian peninsula and North Africa (from Whibley, 2004). (b) The primary pollinators of *Antirrhinum* are solitary bumblebees (*Bombus* spp). Here a bee is seen opening the lower lip of the corolla face to access nectar at the base of the corolla tube and pollen. (Picture from www.bumblebee.org).

Misopates (2n = 16) and sub-shrub *Kickxia* (2n = 18) (reviewed by Vargas *et al*., 2004). Here it is used in its usual sense.

Snapdragon emerged as a model organism during early studies of inheritance and mutation (Darwin, 1868; reviewed by Howard, 2009). *A. majus* was a good model for classical genetics because it has a relatively short generation time of about 4 months, is diploid and can be self- and cross-pollinated (Hudson *et al*., 2008a,b). The potential use of *Antirrhinum* as a molecular genetic model was realised when it became the first plant species from which an autonomous transposon was identified. By cloning the *NIVEA* (*NIV*) gene, Hans Sommer's group was able to identify the autonomous plant transposon, *Tam1*, from an unstable *niv* allele (Bonas *et al.,* 1984). Subsequently, several other transposons were isolated from the *Nivea* locus including *Tam3*, which was in turn used as a molecular tag to isolate *Pallida* (Martin *et al.,* 1985; Sommer *et al.,* 1985). Many other genes, including genes related to flower development, were later cloned from *A. majus* using the transposon tagging system (Coen *et al.*, 1990; Sommer *et al.*, 1985; Bradley *et al.*, 1996).

The genus *Antirrhinum* can also be considered a good model for studies of natural variation (Cubas, Vincent and Coen; 1999; Vieira and Charlesworth, 2002; Scwhinn *et al*., 2006; Whibley *et al*., 2006; Tastard *et al*., 2008), because although its species have diverse morphology, they can be crossed with each other to identify the genes responsible for their differences (Stubbe *et al*., 1966; Whibley *et al*., 2006). However, most of the natural variation is not genetically simple and each trait is likely to involve several genes and gene interaction (such as epistasis), and therefore the correlation of the phenotype to the genotype at any one locus will be low. One approach to identifying genes underlying complex traits is as quantitative trait loci (QTL). This is made possible by the use of a high-density of molecular markers to identify chromosome regions inherited from each parent and to identify correlations between these and the phenotype, measured quantitatively. Molecular recombination maps have been made for hybrids between several *Antirrhinum* species (Schwarz-Sommer *et al*, 2003; Yang, 2007; Greenshields, 2007; Feng *et al*., 2009; Schwarz-

Sommer *et al*, 2010) and been used to identify QTL for complex traits such as flower morphology and leaf size and shape.

1.2 Flower colour and pollinators attraction

Pigment accumulation in specific plant tissues such as anthocyanins and aurones in flowers, pollen and fruits is believed to have evolved as one of the strategies to attract animal pollinators or seed dispersers (Lepiniec *et al*., 2006).

Differences in floral pigmentation have been clearly associated with attraction of different pollinators in many studies. For example, Whittall *et al*. (2006), working with *Aquilegia* found that the species lacking anthocyanins, with white, cream or yellow flowers, were usually pollinated by hawkmoths, compared to anthocyanin-containing species which were generally pollinated by bees and hummingbirds. Bradshaw and Schemske (2003) further provided evidence that pollinators such as bumblebee and hummingbirds can distinguish between different coloured *Mimulus* flowers. *M. cardinalis* flowers are orange-red because they produce both anthocyanins and carotenoids, and are primarily pollinated by hummingbirds. *M. lewisii* has pink flowers because it can produce only anthocyanins and has bumblebee as its primary pollinator. Inter-species hybrids identified a QTL – the *YELLOW UPPER* (*YUP*) locus – responsible for carotenoid deposition. The *YUP* allele from *M. cardinalis* was introgressed into the genetic background of *M. lewisii* by repeated back-crossing to create a near isogenic line (NIL) with orange-red flowers. A NIL was made in the opposite direction to produce a deep pink NIL in the background of *M. cardinalis*. These suggested that the pollinators discriminate flower colour because bumblebee visits were 74 times more frequent to deep-pink *M. cardinalis* than to the wild-type and hummingbird visits 68 time more frequent for orange *M. lewisii* than the wild-type. However, the involvement of other introgressed genes linked to *YUP* could not be ruled out. Hoballah *et al*. (2007) eliminated this possibility in their study of pollinator attraction in *Petunia*. White flowered *P. axillaris*, which has a loss-of-function mutation in the *AN2* gene needed for anthocyanin production, is usually pollinated by hawkmoths (family Sphingidae).

Transgenic *P. axillaris* carring a functional *AN2* allele from the pink-flower *P. integrifolia* had pink flowers, that were visited less by hawkmoths than the white-flowered wild-type, but more by bumblebees, which are the major pollinators of *P. integrifolia*. This result confirmed the role of flower colour in attracting particular pollinator species and that a single gene can have a significant role in changing pollinator discrimination.

The major pollinators of *Antirrhinum* species are large, solitary bees, reflecting their specialised flower structure. A study of three *Antirrhinum* species, *A. graniticum* (subsection *Antirrhinum*), *A. braun-blanquetii* (subsection *Streptosepalum*) and *A. charidemi* (subsection *Kickxiella*) found that seven bee species (*Bombus hortorum*, *Anthidium manicatum*, *Chalicodoma lefebvrei*, *Anthidium sticticum*, *Anthophora dispar*, *Xylocopa violacea* and *Anthidium cingulatum*) were responsible for 90% of visits, though little difference was found in the bee species visiting each *Antirrhinum* species (Vargas *et al*., 2010). Pollination is crucial to the reproductive success of most *Antirrhinum* species, because they have a genetic self-incompatability system that prevents self-fertilization (Vieira and Charlesworth, 2002). A few species, including cultivated *A. majus*, are self-compatible, though outcrossing in *A. majus* produces more seeds per fruit than self-pollination, which can increase fitness (Glover and Martin, 1998).

It has previously been observed that mutations altering flower colour in *A. majus* can affect fitness (Glover and Martin, 1998) and this is related to the way in which bees see flowers (Kevan *et al*., 2001; Vorobyev *et al*., 1997; reviewed by Dyer *et al*., 2007). The colour vision of bumblebees is based on three classes of photoreceptors, the most sensitive in the UV range at 350 nm, in the blue at 440 nm and in the green at 540 nm (Peitsch *et al*., 1992; cited by Dyer *et al*., 2007). The *nivea* (*niv*) mutant of *A. majus* lacks the chalcone synthase enzyme needed for production of both flavonoids and anthocyanins and therefore has flowers that appear albino to the human eye and similar to most white flowers in nature. However, bees are almost unable to detect *niv* mutant flowers because of their altered spectral reflection. A normal white flower does not reflect UV, making it appear coloured to bees (Kevan

et al., 1996). In contrast, a *niv* mutant flower does reflect UV, mostly wavelengths under 360 nm, causing it to look uncoloured to the bee's eye (Dyer *et al.,* 2007). Because bees do not distinguish by brightness, the *niv* mutant flowers do not catch their eyes (Waser and Chittka, 1998). In contrast, wild type *Antirrhinum* flowers have a reflectance peak in the violet near 430 nm, a reflectance minimum in the green/yellow (around 550 nm) and a sudden increase in reflectance in the red, above 600 nm. They therefore appear to be coloured but dark to bees (Dyer *et al*., 2007). The combination of violet and red reflectance makes wild-type *A. majus* look purple to the human eye, but the red component is much less relevant for bees, whose visual spectrum is very limited in the red (Chittka and Waser, 1997).

If bees can detect colour variation in *Antirrhinum*, there is the crucial question of whether they prefer a particular colour and why. Recent studies have shown that bees associate darker colours in *Antirrhinum* flowers with warmth (Dyer *et al*., 2006). Because darker coloured flowers reflect less light, they are usually warmer and bees were found to prefer warmer flowers, whether purple or pink.

The shape of petal epidermal cells has also been shown to be important in – *mixta* mutants, lacking conical epidermal cells are discriminated against by foraging bumblebees (Comba *et al*., 2000; Whitney *et al*., 2009). Petals of the *mixta* mutant, having flat epidermal cells, had previous been shown to reflect more light than wild-type and therefore to appear less intensely pigmented (Noda *et al*., 1994). However, the use of textured, but unpigmented petal replicas showed that conical cells providing a better grip for bumblebees trying to extract nectar or pollen, (Whitney *et al*. 2009) and possibly also make the flower slightly warmer (Comba *et al*., 2000). In conclusion, bees seem likely to discriminate flower colours not because of a preference for colour itself, but because colour is associated with other factors such as warmth and a suitable landing surface that the bees have learnt are beneficial from previous visits. However, studies in the wild have shown that if the pollinators are abundant and the pollination rates are very high, the importance of factors such as flower pigmentation or conical cell shape is diminished (Jones and Reithel, 2001).

Many bee species do not visit differently coloured flowers of the same species at random, but preferentially visit flowers of the same colour, a phenomenon termed constancy. This was demonstrated for *A. majus* in an experiment designed to test discrimination by bees for yellow flowers (carrying an *incolorata* mutation, which prevents anthocyanin production, and the *Sufurea* gene, which allows production of yellow aurone pigments throughout the corolla) compared to wild-type magenta flowers with aurone resticted to the corolla lip. No overall preference was found for magenta or yellow flowers, but individual bees rarely switched between colours (Jones and Reithel, 2001). This study also found that visitation time on one flower colour could be increased by increasing the nectar reward artificially, and that visiting time affected the amount of pollen carried from the flower (i.e., male fitness), but not the number of seeds set (i.e., female fitness). It also demonstrated that the bees were able to discriminate different coloured *Antirrhinum* flowers, although the study was carried out in the USA and involved pollinators that are not native to Europe.

One more aspect of flower colour that is worth mentioning is venation pattern. Venation (contrasting pigmentation over floral veins) is one of the most common patterns in plants and, as shown in field observations recently, can assist in attraction of pollinators (Shang *et al*., 2010). However, it is still unclear how pollinators perceive venation and how venation can attract pollinators. One possibility is that the vein pattern guides pollinators to nectar and pollen, reducing the investment in time needed for the pollinator to gain its reward.

1.3 Flower pigmentation: Anthocyanins

Although there are many colours of flowers, the pigments involved can be divided into only three chemical classes, the carotenoids, the betalanins and the flavonoids. Carotenoids, lipid-soluble pigments giving yellow and orange colours, are the most common but generally contribute less to flower coloration than other pigments. The betalanins, found specifically in Caryophyllales, give red colours. The flavonoids, water soluble pigments, make the major contribution to the colour of flowers in most

Angiosperms: flavonols and flavones give ivory and cream colours; aurones and chalcones give yellow and orange colours and anthocyanins give red-pink-purple and blue colour ranges. These three groups of flavanoids have been studied intensively, and, at the molecular genetic level, anthocyanin biosynthesis is more clearly understood than the others (Whitney and Glover, 2007). However, the regulation of anthocyanin production still remains unclear (Schwinn *et al*., 2006).

1.3.1 Anthocyanin biosynthesis

Seven core structural genes are known to encode enzymes that catalyze synthesis of anthocyanins. The pathway of anthocyanin biosynthesis in *Antirrhinum* is shown in Figure 2.

Most of the genes encoding enzymes of the anthocyanin biosynthetic pathway in *Antirrhinum* have been characterized in the last three decades. Although regulators of the pathway have also been identified, others remain obscure. In this introduction, I will summarize the literature on anthocyanin biosynthesis enzymes and genes in three major model species for flower corolla pigmentation, Petunia (*P. hybrida*), morning glory (*Ipomoea* species) and *Antirrhinum*.

i. Chalcone Synthase

Chalcone synthase (CHS) catalyzes the first step in the biosynthesis of flavonoid and anthocyanin pigments. It is one of the best-characterized enzymes of plant secondary metabolism in term of biochemistry (Springob *et al*., 2003). CHS was classified as one of the type III polyketide synthases (PKSs) (Austin and Noel, 2003). It performs sequential condensations with *ρ*-coumaroyl-CoA and three acetate units from malonyl-CoA. The result is naringenin chalcone (2',4,4',6'-tetrahydroxychalcone), which in plants, can be stereospecifically converted to flavonone *(2S)*-naringenin by chalcone isomerase (CHI) (reviewed by Springob *et al*., 2003). Despite a large number of predicted CHS sequence in the Genbank database, the high sequence homology to other plant PKSs make it impossible to predict activity from only sequence data and therefore several predicted CHS may have different metabolic

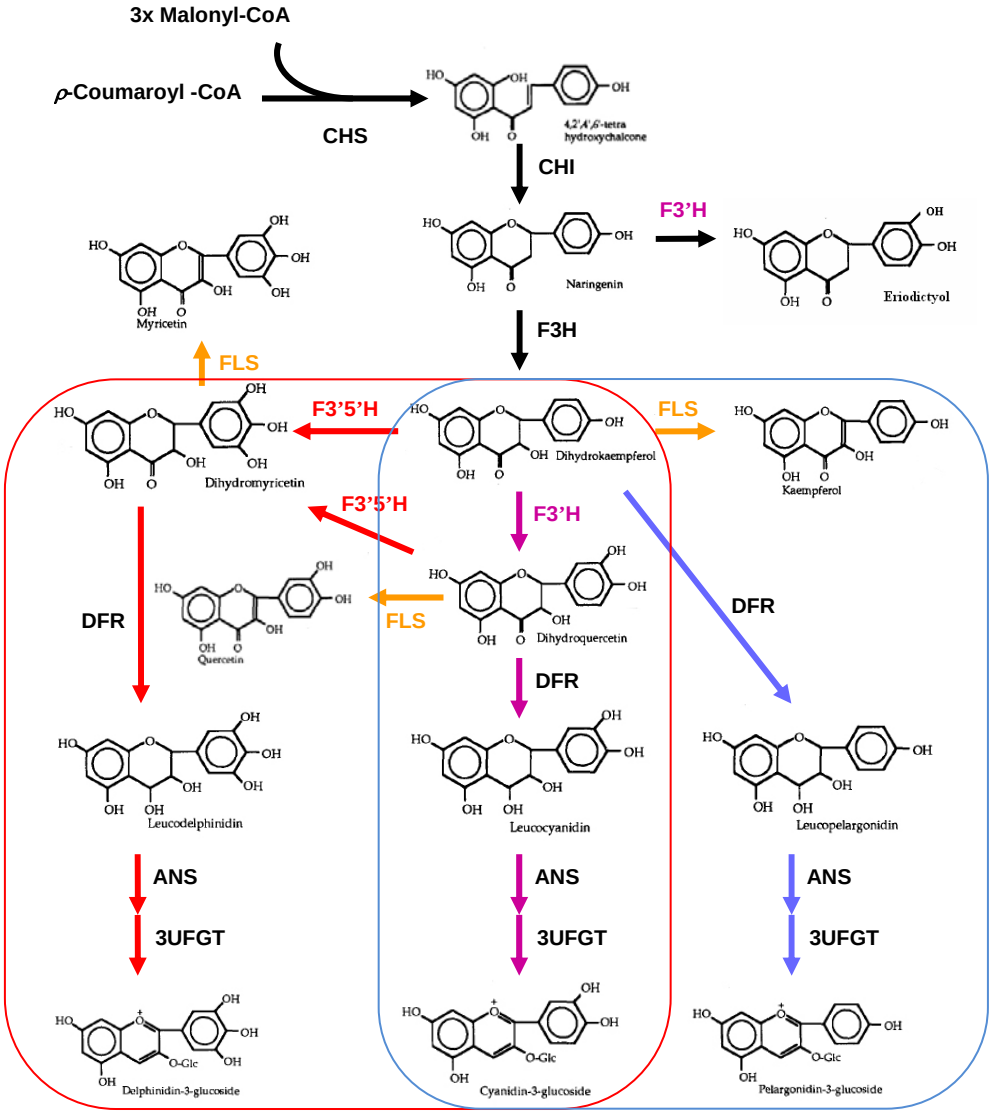


Figure 2. The anthocaynin biosynthesis pathway in *Antirrhinum* and *Petunia.* (Modified from Holton and Cornish, 2005).CHS, chalcone synthase; CHI, chalcone isomerase; F3H, flavonol - 3-hydroxylase; F3'H, flavonoid - 3' –hydroxylase, F3'5'H , flavonoid - 3'5' –hydroxylase, FLS, flavonol synthase, DFR, dihydroflavonol reductase; ANS, anthocyanidin synthase; UF3GT, UDP glucose flavonoid-3-glucosyltransferase.The cyanidin branch of the pathway is shown to the right, the pelargonidin branch is in the middle and the delphinidin branch is in the left. Cyanidin and Delphinidin differs from Pelargonidin by possessing an extra hydroxyl group at 3' position and 3', 5' position, which is added by F3'H and F3'5'H, respectively. This enzyme creates a branch in the biosynthetic pathway. *Antirrhinum* only produces cyanidin- and pelargonidin-type of anthocyanin whereas *Petunia* only produces cyanidin- and delphinidin-type of anthocyanin because DFR in *Petunia* inefficiently reduces dihydrokaempferol (Johnson *et al*., 2001).

roles, which can be only elucidated by experimental characterization.

Sequence data predict that many plant species contain small *CHS* multigene families. Most *CHS* genes studied contain a single intron at a conserved site, with the exception of *A. majus*, which has a second intron (Sommer and Saedler, 1986). In *Petunia*, a small family of at least 12 *CHS* genes were found, though only two genes (*CHS-A, CHS-J*) were actively transcribed in flower tissue. However, the mRNA level of *CHS-J* is approximately 10 times lower than that of *CHS-A*, suggesting that *CHS-A* has a predominant role in floral anthocyanin production (Koes *et al*.,1986; Koes *et al*., 1989). In Morning Glory, six functional *CHS* genes were found. In the same manner, only one (*CHS-D*) was found to be responsible for anthocyanin production in the corolla limb (Durbin *et al*., 2000a,b) whereas *CHS-E* was believed to be involved in pigment production in the corolla tube (Johzuka-Hisatomi *et al*., 1999). In *Antirrhinum*, at least two functional *CHS* genes have been reported so far (Martin *et al*., 1991; Hatayama *et al*., 2006). Transcripts from the *CHS* that had been most intensively studied, *Nivea* (*Niv*) or *AmCHS1*, specifically accumulate in petals. The second gene (*AmCHS2*) gene was isolated from yellow flowered *Antirrhinum* and is expressed at a lower level than *Niv* (Hatayama *et al*., 2006). Null niv^- alleles are recessive to Niv^+, and lead to production of albino flowers lacking all flavonoids. However, several semi-dominant *niv* alleles (*niv-571*, *niv-572* and *niv-525*) give very pale flowers when heterozygous with Niv^+. The semi-dominant alleles appear to acts in *trans* to inhibit expression of their *Niv* homolog. Molecular characterization of one of the semi-dominant alleles found that it contains multiple truncated copies, one of which is in an inverse orientation, which authors suggested might generate inhibitory antisense RNA, though this was not detected (Bollmann, Carpenter and Coen, 1991). More recent understanding of epigenetics suggests that their effect might involve siRNA, though this was not tested at the time.

ii. CHI (Chalcone isomerase)

Chalcone isomerase (CHI) catalyzes the stereospecific cyclization of naringenin chalcone (2′,4,4′,6′-tetrahydroxychalcone) and 6′-deoxychalcone (2′,4,4′-trihydroxychalcone) to (2S)-naringenin (4′,5,7-trihydroxyflavanone) and (2S)-

liquiritigenin (4′,7-dihydroxyflavanone), respectively. CHIs (CHI-type I enzymes) of non-legumes convert only chalcones which are hydroxylated in the 6′-position. In legumes, the type II CHI was found to convert both 6'-hydroxy- and 6'-deoxychalcones (Forkmann and Heller, 1999; Dixon *et al*., 1988: reviewed by Spirngob *et al*., 2003). It has been suggested that legume-type *CHIs* evolved from an ancestral common-type *CHI* by gene duplication (Shimada *et al*., 2003; Ralston *et al*., 2005). Even in the absence of CHI, chalcones with free 2'- and 6'-hydroxyl groups can spontaneously isomerize to flavanones, however, this occurs at a much slower rate than the CHI-catalyzed reaction and yields two stereoisomers, (*2R*)-flavanones and (*2S*)-flavanones (Stafford, 1990). Since (*2S*)-flavanones are the only substrates of downstream enzymes of the flavonoid pathway *in vivo*, CHI is required for the efficient formation of biologically active (*2S*)-isomers (Springob *et al*., 2003). This assumption was confirmed by *chi* mutants in Arabidopsis; mutants lacking CHI activity, such as the Arabidopsis *transparent testa 5* (*tt5*) mutant, accumulate only trace amounts of flavonoids (Shirley *et al*., 1995).

CHI-like sequences are unique to higher plants and show no similarity with other proteins (Springob *et al*., 2003). Like CHS, small families of *CHI* genes are found in some higher plant species. In Petunia, two functional copies of *CHI* were isolated; *Chi-A* (the *PO* locus), which was found to be expressed in flower tissue and seedling and induced by light and UV stress, and *CHI-B*, which is restricted to immature anthers and pollen (van Tunen *et al*., 1988, van Tunen *et al*., 1990). Petunia CHI-A was found to be active in converting only 2′,4,4′,6′- tetrahydroxychalcone (Dixon *et al*., 1988). In contrast, only a single copy of *CHI* has so far been found in morning glory and *Antirrhinum* (Clegg and Durbin, 2003; Martin *et al*., 1991).

iii. Flavanone-3-hydroxylase

Flavanone-3-hydroxylase (F3H) is classified as a soluble 2-oxoglutarate-dependent dioxygenase, according to its requirement for the cofactors 2-oxoglutarate, oxygen, ferrous (Fe^{2+}) ions and ascorbate. F3H converts the flavonones ([2s]-naringenin and [2S]-eriodictyol) to dihydroflavonols ([2R, 3R]-dihydrokaempferol and [2R, 3R]-

dihydroquercetin) by a hydroxylation in position 3 (reviewed by Springob *et al*., 2003).

Genetic studies in Petunia revealed that the *AN3* (*ANTHOCYANIN3*) *l*ocus encodes flavanones-3-hydroxylase (Froemel *et al*. 1985; Britsch *et al*. 1992). In morning glories, F3H is encoded by at least two genes in the *I. purpurea* genome, both of which are expressed (McCaig, 1998: reviewed by Clegg and Durbin, 2000), whereas one functional copy of F3H and one pseudogene had so far been identified in *I. nil* (Hoshino *et al*. 1997). In *Antirrhinum*, only one single F3H gene (*Incolorata*) has been found so far. This gene is considered likely to contribute flower colour variation in *Antirrhinum*, because *Inc* wild-type alleles are not fully dominant over loss-of-function mutations (Stubbe, 1966), therefore F3H was believed to catalyse the rate-limiting step of anthocyanin biosynthesis. Its potential involvement in natural variation was further supported by allelism tests, in which different *Antirrhinum* species were crossed to wild-type and *inc* mutant *A. majus*, although the results of these tests were not definitive because other loci and genetic interactions are likely to be involved (Chaffe, 2002). Similarly, QTL analysis had suggested that *Inc* could be involved in the differences between *A. majus* (magenta flowers) and *A. molle* (pale flowers) (Greenshields, 2007). In contrast, Whibley (2004) found that variation in *Inc* is unlikely to have a role in the flower colour differences between *A. majus* subspecies. *pseudomajus* (magenta flowers) and *striatum* (lacking anthocyanin).

iv. Flavonoid-3'-hydroxylase and Flavonoid- 3',5'-hydroxylase

Flavonoid-3'-hydroxylase (F3'H) and Flavonoid- 3',5'-hydroxylase (F3'5'H) belong to the CYP75B- and CYP75A P450-monooxygenase families, respectively. These two enzymes are responsible for flavonoid B-ring hydroxylation and are required for biosynthesis of flavones, flavanones, flavonols and anthocyanin. In the case of anthocyanins, F3'H is needed for the formation of cyanidins, and F3'5'H for the delphinidins (reviewed by Springob *et al*., 2003).

In flowers of Petunia, *HT1* and *HT2* encode F3'H activity, while *HF1* and *HF2* encode F3'5'H activity. *HT1* and *HF1* are expressed in both corolla limb and tube, whereas *HT2* only affects the tube and *HF2* expression was found only in the limb. These four genes show complex epistatic relations (review by Tornielli, Koes and Quattrocchio, 2009).

An additional Petunia gene, encoding a cytochrome b5 (CYTb5), is also required for 3'5' hydroxyl addition of anthocyanins (Brugliera *et al*., 2000). This gene was first named *DIF-F* (*DIFFERENTIAL-F)* and its transcripts accumulate in the limbs and tubes of the flower corolla and in ovaries, but not in vegetative tissue. (de Vetten *et al*., 1999)

F3'H has been characterized in *I. purpurea, I. tricolor* and *I. nil* (Hoshino *et al*., 2003), whereas the dihydroflavonol 3'5'-hydroxylase (F3'5'H) has not yet been characterized in *Ipomoea*. *Antirrhinum* produces only F3'H and a gene encoding the enzyme had been reported (reviewed by Holton and Cornish, 1995)

v. Dihydroflavonol-4-reductase

Dihydroflavonol-4-reductase reduces (2R,3R)-dihydroflavonols (dihydrokaempferol, dihydroquercetin and dihydromyricetin) to (2R, 3S, 4S)-leucoanthocyanidins (leucopelargonidin, leucocyanidin and leucodelphinidin) (Springob *et al*., 2003).

Petunia DFR is well known to have a very specific substrate. It is unable to accept dihydrokempferol as a substrate and therefore petunia cannot produce pelargonidin-type anthocyanins. It was confirmed using chimeric DFR proteins expressed from *Petunia* and *Gerbera* genes, that one single amino acid substitution could change the substrate specificity, enabling petunia to produce pelargonidin-type compounds (Johnson *et al*., 2001). Petunia contains three different *DFR* genes (*DFR-A*, *DFR-B*, *DFR-C*), but only the *DFR-A* (*An6*) gene is transcribed in floral tissues (Beld *et al*., 1989). In morning glory, at least three copies of *DFR* genes (*DFR-A*, *DFR-B* and *DFR-C*) are present (Inagaki *et al*., 1999). *DFR-B* had been identified as the gene responsible for anthocyanin production in the corolla limb, based on the finding that

a *dfr-b* mutation carrying a large transposon in the second intron, resulted in loss of pigment from the limb of *I. nil* (Inagaki *et al*., 1994).
In *Antirrhinum*, only one gene had been found to encode DFR, called *Pallida* (Martin *et al*., 1991)

vi. Anthocyanin synthase

Anthocyanin synthase (ANS) is involved in a crucial step of anthocyanin formation, because it catalyses the oxidation of colourless leucoanthocyanidin to the precursor of anthocyanidin. The enzyme belongs to the 2-oxoglutarate-dependent dioxygenase enzyme family, like flavanone-3-hydroxylase (F3H) and flavonols synthase (FLS) (Springob *et al*., 2003). Biochemical activity of ANS was confirmed *in vitro* in 1999 using *Perilla frutescens* ANS expressed in *E. coli*. The purified ANS showed 2-oxoglutarate-dependent-oxygenase activity, catalyzing the oxidation of leucoanthocyanidin to anthocyanidin in the presence of ferrous iron, 2-oxoglutarate and ascorabate, following acidification with HCl (Saito *et al*., 1999).

A Petunia gene (*AN17*) encoding ANS, had been isolated and functionally characterized (Weiss *et al*., 1993; Nakajima *et al*. 2001).

The *Antirrhinum* gene (*Candica*) encoding ANS, was cloned by Martin *et al*. in 1991 but is not yet well characterized.

vii. Anthocyanin glycosylation

In the majority of plants, glycosylation of a hydroxyl group at the C-3 position of anthocyanidin is the primary modification step needed to stabilize anthocyanidin and is required for further modifications, such as secondary glycosylation, acylation, methylation and prenylation. UDP-glucose:flavonoid-3-0-glucosyl transferase (3GT) glycosylates anthocyanidins and flavonols on the 3 position, using flavonoid as the sugar acceptor and UDP-sugars as the sugar donor, resulting in anthocyanin-3-glucoside and flavonols-3-glucoside, respectively (Springob *et al*., 2003; Yonekura-Sakakibara *et al*., 2009).

Genes encoding 3GT have been characterized in many species. In *P. hybrida*, a *3GT* gene called *PGT8* was isolated by Yamazaki *et al*. (2000), who also found evidence for a second gene in the *Petunia* genome. In *Antirrhinum*, one *3GT* gene has been cloned but is still not well characterized (Martin *et al*., 1991). In morning glories, a single copy *3GT* gene, called *3GGT*, has been found in *I. purpurea* and *I. nil* (Clegg and Durbin, 2000; Morita *et al*., 2005).

The 3-glucosyl anthocyanin can be then further modified by addition of a rhamnose molecule to the glucose at position C-3, creating anthocyanin-3-rutinoside, catalysed by 3-rhamnosyltransferase. The corresponding gene was isolated from Petunia (*RT*; Brugliera, *et al*., 1994, Kroon *et al*., 1994) and has also been functionally characterized in Arabidopsis (*UGT78D1*; Jones *et al*., 2003). However, although anthocyanin-3-rustinosides are found in *Antirrhinum* and morning glory, isolation of a gene encoding the rhamnosyltransferase has not been reported so far. In *I. nil,* a *3GGT* gene was found to be similar to the Petunia *Rt* gene, but its product found to show only 3GT activity (Morita *et al*., 2005).

Glucosylation at the 5-O-position of anthocyanins, which is catalyzed by anthocyanin -5-O-glucosyltransferase (5GT), is the other modification step that is commonly found in flowering plants. In Petunia, the *PH1* gene was reported to encode a 5GT which is uniquely strict in its substrate requirement, being specific to delphinidin-3-(ρ-coumaroyl)-rutinoside indicating an earlier acylation step at the C-3 position (Jonsson *et al*., 1984a, Yamazaki *et al*., 2002). In contrast, *Antirrhinum* appears not to have the 5-O-glucosyltransferase activity because it produces only cyanidin- and pelargonidin-3-rutinosides in flowers (Harborne, 1963). Similarly, there has been no report of *5GT* gene isolation from *Ipomoea*, although *I. purpurea* flowers accumulate anthocyanins that have been glycosylated at the 5 position, such as cyanidin- or pelargonidin- 3-(2-glucosylcaffeoylglucosyl-6-caffeoylglucosylcaffeyl-glucoside)-5-glucoside, -3-(2-caffeoylglucosyl-6-caffeoyl-glucoside)-5- glucoside, -3-(2-glucosylcaffeoylglucosyl-6-caffeoyl-glucoside)-5-glucoside (Lu *et al*., 2009.).

viii. Further anthocyanin modification: acylation and methylation

Acylation is one of the most common modifications of plant secondary metabolites including anthocyanins and more than 65% of known anthocyanins appear to be acylated (Anderson and Jorheim, 2006; reviewed by Yonekura-Sakakibara *et al*., 2009). The acylated anthocyanins can be divided into two groups, depending on whether the group added to the glycosyl moieties of anthocyanin are aromatic or aliphatic. The most common aromatic additions are hydrozycinnamoyl groups, such as *ρ*-coumaryl, caffeyl, etc. and malonyl is the most common aliphatic group, though acetyl, succinyl and others are also found (Springob *et al*., 2003; Yonekura-Sakakibara *et al*., 2009). The first gene encoding an enzyme that catalyzes the transfer of aromatic acyl groups to anthocyanins was isolated from gentian, *i.e.*, encoding hydroxycinnamoyl CoA: anthocyanin 5-glucoside hydroxycinnamoyl CoA transferase (5AT) activity (Fujiwara *et al*., 1998).

Methylation has been reported for about 20% of anthocyanidins, including peonidin, petunidin and malvinidin. An enzyme catalyzing anthocyanin methylation has been found to belong to the Class I S-adenosyl-L-methinonine (SAM) dependent-O-methyltransferases (OMTs) (reviewed by Yonekura-Sakakibara *et al*., 2009).

Although Petunia had been found to have further modifications to anthocyanin, there have been no reports of acylation or methylation of anthocyanins in *Antirrhinum* and morning glory. However, modification involving glycosylation and/or acylation have been reported in roses, sweet peas and gentians, which are of interest because of the ornamental value of these species (Otaga *et al*., 2005; Noda *et al*., 2004; Fukuchi-Mizutani *et al*., 2007; reviewed by Tanaka, Brugliera and Chandler, 2009).

Though methylated anthocyanin has not been reported in snapdragon, methylation of other secondary plant products in their flowers is involved in the production of scent, such as methylbenzoate. In cultivated *A. majus* this is produced by S-adenosyl-L-methionine:benzoic acid carboxyl methytransferase (BAMT), and the gene encoding this activity has been cloned (Kolosova *et al*., 2001). The enzyme and homologous

gene have been found in Petunia which also produce methylbenzoate (Negre *et al*.,2003).

In Petunia, genes required for the modification steps following the first glycosylation have been described; the *GF* gene encodes the enzyme that attaches either caffeic acid or coumaric acid to the anthocyanin-3-rutinoside (Jonsson *et al*., 1984b). However, only delphinidin-3-(ρ-coumaroyl)-rutinoside will be converted into the corresponding 5-O-glucoside by UDP:glucose anthocyanin -5-O-glucosyltransferase, encoded by the *PH1* genes which are known to have at least two copies (Yamazaki *et al*., 2002). Finally methylation can occur, catalysed by anthocyanin 3'-OMT and anthocyanin 3',5'-OMT, which are encoded by genes, *MT1*/*MT2* and *MT1*/*MT2*, respectively. The predominant methylation in Petunia involves delphinidin derivatives, producing petunidin- and malvidin-glucoside by the action of anthocyanin 3'-OMT (Yonekura-Sakakibara *et al*., 2009).

1.3.2 Transport and accumulation of anthocyanin

Anthocyanins biosynthesis has been localized to the cytosol, near the endoplasmic recticulum membrane system (Sun, Li and Huang, 2011). However, the stability of anthocyanins in aqueous solution with near neutral pH, as in the cytosol, is low (Brouillard, 1982; Yonekura-Sakakibara *et al*., 2009). For this reason, anthocyanins need to be transported and stored in vacuoles, which have lower pH, for prolonged accumulation. Although the biosynthesis of anthocyanin has been studied intensively for several decades, little is known about vacuolar sequestration of anthocyanin. However, several mechanisms have been investigated so far.

i. pH-dependent transport system

There have been several reports of pH-dependent accumulation of flavonoids in vacuoles, such as apigenin -7-(6-O-malonyl)-glucoside in parsley (Matern, Heller and Himmelspach, 1983) and sinapoylated anthocyanin in carrot cell suspension (Hopp and Seitz, 1987; cited by Vogelien *et al*., 1990) or evidence of apigenin 6-*C*-glucosyl-7-*O*-glucoside uptake in barley mesophyll (Klein *et al*.,1996; cited by

Frangne *et al*., 2002). However, in this pH-dependent transport system, acylation of flavonoid substrates appears to be essential (Springkob *et al*., 2003)

ii. Anthocyanin trafficking system: Glutathione-*S*-transferase acts as the carrier for ABC transporter and MATE transporter

Two proposed anthocyanin trafficking mechanisms require the involvement of Glutathione-*S*-transferase (GST). The first model, the transporter-mediated model (Zhoa and Dixon, 2010), involves ABC-type transporters. While the second model, the vesicle trafficking-mediated model, suggests the involvement of vacuole-localized multidrug and toxic compound extrusion (MATE) transporters (Grotewold and Davis, 2008).

The relationship between GST and anthocyanin transport is not yet clear. Glutathione-modified anthocyanin molecules have not been found in plant cells, suggesting that GST does not act on anthocyanins. However, several additional lines of evidence offer clues. In Petunia, the *AN9* gene, required for accumulation, encodes GST, but is unable to conjugate glutathione to anthocyanin or other flavonoid molecules *in vitro* (Mueller *et al*., 2000). A recent study has provided evidence that *TT19*-encoded GST in Arabidopsis might act as a carrier to transport anthocyanin from the cytosol to the tonoplast. This study showed that *TT19*-encoded GST has GST activity; but was unable to use cyanidin and cyainidin-3-glucoside as substrate, though it was needed for their accumulation (Sun, Li and Huang, 2011). However, it did directly bind to the cyanidin and cyanidin-3-glucoside *in vitro* and increased the water solubility of cyanidin. Consistent with the proposed role of TT19 as a carrier, it was found to be located not only in the cytosol but also on the tonoplast membrane. Therefore the pigment-TT19 complex might form in the cytosol and be released to the vacuole at the tonoplast. Because acylation is needed for accumulation in the vacuole, release or uptake at the tonoplast might involve acylation.

- **Involvement of an ABC-type transporter**

Evidence suggesting the involvement of ABC-type transporter in anthocyanin accumulation was reported by Goodman and his colleagues (Goodman *et al*., 2004). Maize plants expressing antisense RNA for *ZmMRP3*, which encodes an ABC transporter, showed a reduction in anthocyanin production and mislocalization of pigment in the cytosol (Goodman *et al*., 2004). It is therefore possible that the ABC-transporter may bind anthocyanins-GST complex through hydrophobic interactions and then escort them to the tonoplast membrane.

- **Involvement of MATE family transporters**

This model was first proposed based after the observation that anthocyanins are accumulated first in vesicle-like structures, called anthocyanoplasts, and then maybe imported into the vacuole probably in a vesicle-like manner. (Grotewold *et al*., 1998; Gomez *et al*., 2011). The involvement of MATE transporter in this model is based partly on the finding that a pH gradient across the anthocyanoplast/tonoplast membrane, generated by V-ATPase and vacuolar H^+-pyrophosphatase, is needed for flavonoid uptake (Klein *et al*., 1996), as MATE transporters use secondary transport mechanism depending on proton gradient across the anthocyanoplast/vacuolar membrane to function by mechanism of flavonoid/ H^+ exchange (Yazaki, 2005). Then, the GST might act as a carrier and the anthocyanin-GST complex using MATE transporters to transport anthocyanin. Several lines of evidence suggest that MATE family members are involved in anthocyanin/proanthocyanidin accumulation. For example, the product of the tomato *MTP77* gene (later called putative anthocyanin permease, *PAT*) was up-regulated in response to over-expression of a MYB transcription factor that promotes anthocyanin synthesis and accumulation (Mathew *et al*., 2003; Butelli *et al*. 2008). The *TT12* gene of Arabidopsis, required for anthocyanin accumulation in the seed coat, encodes a MATE transporter that has been reported to carry cyanidin-3-O-glucoside. The related MATE1 protein in *Medicago truncatula* can transport epicatechin -3'-O glucoside and cyanidin-3-O-glucoside (Zhoa and Dixon, 2009) while MATE2 from the same species can transport anthocyanins (pelarogonidin- and cyanidin-3-O-glucoside) and flavone glucosides (naringenin- and luteolin -7-O-glucoside) (Zhoa *et al*., 2011). Other

reports suggest that some tonoplast-localized transporters have a higher activity towards acylated flavonoids (Gomez *et al*., 2009, Zhoa *et al*., 2011), which can explain the requirement for acylation in uptake.

ABC and MATE transporters associated with anthocyanin transport have not yet to be identified in *Antirrhinum*, *Ipomoea* or even Petunia. Although one V-ATPase proton pump gene (*PH5*) has been identified in Petunia showing the involvement in anthocyanin pigmentation, it was clearly shown to be involved only with pH change in vacuole. Loss of PH5 activity decreases the pH of the vacuole lumen, resulting in a flower colour change (Verweij *et al*., 2008).

1.3.3 The regulation of anthocyanin production

Work in several model plant species has identified a complex system involving WD-40, MYB and bHLH transcription factors that regulates anthocyanin biosynthesis in flowers. This system is reviewed mainly with evidence from Petunia and *Antirrhinum*, in which it has been best studied.

i. bHLH proteins

bHLH (basic helix-loop-helix) proteins, also known as MYC proteins, are one of the main transcriptional regulators of the flavonoid biosynthetic pathway. bHLH proteins act as transcription factors and are found in fungi, humans and plants (Massari and Murre, 2000; Pires and Dolan, 2010). In plants, more than 100 bHLH are present in Arabidopsis, and in rice and grape, making it the second largest plant transcription factor gene family after the *MYB* genes (Bailey *et al*., 2003; Li *et al*., 2006). Recent analysis of the plant bHLHs placed them into 26 subgroups (Pires and Dolan, 2010). The bHLHs that control flavonoid biosynthesis fall into one clade - the subgroup IIIf (Pires and Dolan, 2010; Hichri *et al*., 2011).

The bHLH domain consists of about 60 amino acids, of which 19 are highly conserved; about five in the basic region, five in the first helix, one in the loop, and finally eight amino acids in the second helix. Most bHLH proteins bind to DNA as dimers, with the basic helices interacting with adjacent turns of the DNA and the

second helix stabilizing the dimer. About 20% of bHLH proteins in *Arabidopsis* lack the basic region and act as transcriptional repressor by forming heterodimers with a complete bHLH, preventing it from binding with target DNA (Toledo-Ortiz, Huq and Quail, 2003).

The first *bHLH* genes controlling flavonoid biosynthesis were isolated from maize, *B* (*Booster 1*) and *R* (*Red 1*) and therefore the family to which they belong is also called RB (Chandler *et al*., 1989). Member of this subfamily of bHLHs that control flavonoid biosynthesis show several common features, such as the presence of a conserved N-termimal region that interacts with MYB proteins, a negatively charged region necessary to interact with WD40-repeat proteins and/or the RNA polymerase II complex. Within the conserved bHLH domain itself, the C-terminal region is known to be involved in homodimer or heterodimer formation with other bHLHs (reviewed by Hichri *et al*., 2011). Some bHLHs form homodimers and bind the G-box (CACGTGG) or the hexanucleotide E-box (CANNTG), these include the products of the *CrMYC1* (*Cantharanthus roseus MYC1*; Chatel *et al*., 2003), *Perilla MYC-RP* (Gong *et al*., 1999) and *Delila* in *Antirrhinum* (Gong *et al*., 1999). Others can form heterodimers with MYB proteins including Petunia bHLHs, AN1 and JAF13, and Arabidopsis TT8, and the complexs can bind other specific targets (Shimada, Otsuki and Sakuta, 2006).

bHLH genes isolated from petunia so far include *AN1* (the ortholog of *TT8* in Arabidopsis) and *JAF13* (the ortholog of *Delila* in *Antirrhinum* and *GL3/EGL3* from Arabidopsis) (Quattrocchio *et al*., 1998). Although these two genes are both bHLH proteins, they appear to have different evolutionary origin (Spelt *et al*., 2000; Quattrocchio *et al*., 1998). Both AN1 and JAF13 share the ability to interact with similar MYB proteins, such as AN2. However, their functions are not fully interchangeable, since an *an1* mutant, which still has full JAF13 activity, shows a lack of pigmentation in petals and seedcoat cells (Quattrocchio *et al*., 1998; Tornielli, Koes and Quattrocchio, 2009). Though the effect of *JAF13* was proposed to be specific to anthocyanin regulation, the complete picture of its role in anthocyanin biosynthesis is still unclear.

In *Antirrhinum*, the first isolated *bHLH* gene involved in pigmentation was *Delila* (*Del*) characterized by Goodrich *et al*. (1992). A second bHLH gene was subsequently isolated by Schwinn *et al.* (2001) called *Mutabilis* (*Mut*), has not been studied in as much detail. Expression of the two genes differs in *Antirrhinum* flowers. *Del* is active in both petal lobes and the petal tube, while *Mut* is active only in the lobes. The two genes function redundantly to promote anthocyanin synthesis, so that loss of Del activity results in loss of anthocyanin from the petal tube, but not the petal lobes where *Mut* is still active, whereas additional loss of *Mut* activity results in a reduction in colour throughout the flower (Stubbe, 1966; Schwinn *et al*., 2006).

ii. MYB proteins

Plant MYB transcription factors are characterized by the N-terminal MYB domain, consisting of between 1 and 3 imperfect repeats of about 52 amino acids (R1, R2, and R3). The MYB domain is involved in DNA binding and dimerization, while the C-terminal region is involved in regulating target gene expression (i.e. activation or repression) through interaction with other proteins (reviewed by Feller *et al*., 2011 and Hichri *et al*., 2011).

The largest group of plant MYB genes encode two MYB repeats (R2 & R3). Several hundred members of this family are typically found in higher plants. They are highly conserved at the R2R3 domain, but the C-terminal can be variable. The C-terminal region often contains transcriptional activation or repression domains and conserved serine and threonine residues, which may correspond to post-translational modification sites (Martin and Paz-Ares, 1997). Recurring amino acid motifs in the C-termini have been used to cluster the large numbers of *Arabidopsis* R2R3-MYB family into 22 sub-groups (Dubos *et al*., 2010) and can also be applied to R2R3-MYB proteins from other plants, though some differences have been found (Stracke *et al*., 2001; Wilkins *et al.,* 2009). MYB proteins involved in the regulation of flavonoid biosynthesis mostly fall into subgroups 1-7 (reviewed by Feller *et al*., 2011 and Hichri *et al*., 2011).

Most of the MYB transcription factors characterized to date control only one branch of the flavonoid pathway, that leading to anthocyanin. However, recent studies have identified roles for other MYB proteins in regulation of the early step of flavonoid biosynthesis that are shared between flavonols, flavone and anthocyanin (Stracke *et al*., 2007, Zhoa *et al*., 2011).

In Petunia, several R2R3 MYBs were found related to anthocyanin and flower pigmentatation. The *AN2* gene appears to promote pigmentation in petal limbs, because it shows tissue specific-expression (Quattrocchio *et al*. 1999). In other tissues, synthesis is activated by closely related *MYB* genes with different patterns of expression, for example *AN4* in petal tubes and anthers (Kroon, 2004; cited by Koes, Verweij and Quattrocchio, 2005). *PH4* encodes the R2R3 MYB protein that has been associated with acidification of vacuolar lumen but not anthocyanin production (Quattrocchio *et al*., 2006). A closely related gene to *AN4*, previously called *MYBb* and now called *MP*, was isolated from *P. hybrida* x *P. axillaris* hybrids. It is expressed in petals, but its function remains to be characterized (Kroon, 2004; cited by Koes, Verweij and Quattrocchio, 2005). However, Albert *et al*. (2011) demonstrated that *PURPLE HAZE (PHZ)*, one of the new isolated MYB in *P. hybrida* which is closely related to *AN2*, involved in light-induction of anthocyanin pigmentation in vegetative tissue, is the polymorphic allele of *MYBb.* The over-expression of the new MYBs, *DEEP PURPLE (DPL)* and *PHZ* by Albert *et al*. (2011) showed that though they are mainly expressed in vegetative tissue, they can partly complement lack of AN2 activity in flowers, allowing anthocyanin production during flower bud development.

Some MYB proteins not only form heterodimers with bHLH partners, but can also activate expression of the gene encoding the bHLH, providing positive feedback in regulation. This phenomena has been observed for the Petunia AN2 and for the Arabidopisis AtTT2, AtPAP1 and AtPAP2 (Nesi *et al*., 2002, Quattrocchio *et al*., 1998; Borevitz *et al*., 2000; review by Koes, Verweij and Quattrocchio, 2005).

MYB proteins can also function as transcriptional repressors. In Petunia a short R3-MYB, encoded by the *MYBx* gene, can abolish the formation of an activating MYB-bHLH-WD (MBW) complex by sequestering the bHLH, *AN1*, into an inactive complex (Kroon 2004, reviewed by Albert *et al*., 2011). Albert *et al*. (2011) also identified another R2R3 MYB protein in Petunia, PhMYB27, which has high amino acid sequence similarity to strawberry FaMYB1, including presence of a putative ERF-repressor (EAR) domain (Aharoni *et al*., 2001) that was repressed by light, thus, promote anthocyanin production.

The *MYB* genes known to be involved in anthocyanin regulation in *Antirrhinum* so far consists of *Rosea1*, *Rosea2* and *Venosa* (*Ros1*, *Ros2* and *Ve*, respectively; reviewed by Schwinn *et al*., 2006). The combined action of these genes with *Del* and *Mut* provides for complex control of the spatial and temporal production of anthocyanins in the petals.

Ros1 and *Ros2* are tandemly duplicated MYB genes that are expressed throughout petals. Loss of *Ros1* activity reduces anthocyanin pigmentation in *A. majus*, suggesting that *Ros1* promotes anthocyanin expression. The function of *Ros2* is less clear because the gene is partially deleted in *A. majus*, which has darkly pigmented flowers, but active in some other *Antirrhinum* species with paler flowers. One possibility is that *Ros2* is a repressor of anthocyanin synthesis and that its deletion in *A. majus* leads to an increase in pigmentation. However, *Ros2* can activate anthocyanin production when expressed from the *35S* promoter in *ros1* mutant petal cells, supporting an alternative role as an activator. *Venosa* (*Ve*) encodes a related MYB protein that promotes pigmentation only in epidermal cells overlying veins, producing a venation pattern in plants with reduced *Ros1* activity (reviewed by Schwinn *et al*., 2006). The different action of the genes is therefore partly the result of cell-specificity of expression.

iii. WD-40 repeat proteins

WD40 proteins contain 4-16 copies of highly conserved repeating units, of ~40 amino acids, which usually end with Tryptophan and Aspartic acid (WD) residues

(reviewed by Smith *et al*., 1999). This motif can be tandemly repeated for 4-16 times within the protein. In Arabidopsis, there are an estimated 239 proteins with at least four WD motifs. WD40 proteins are found in all eukaryotes and are implicated in many kinds of functions, including RNA processing, signal transduction, cytoskeleton assembly and cell division (reviewed by Smith *et al*., 1999). WD40 transcription proteins appear unable to bind directly to DNA, however, they can interact with several proteins simultaneously (van Nocker and Ludwig, 2003). The first WD40 protein found to be required for anthocyanin pigmentation was AN11 in petunia (de Vetten *et al*. 1997).

AN11 in petunia had been found to be expressed throughout the plant, suggesting that WD40 is not involved in tissue specificity of pigmentation. It appears to be necessary for anthocyanin production in all tissues, because *an11* mutants completely lack anthocyanin (de Vetten *et al*., 1997).

Screening of a floral cDNA libraries from *Ipomoea nil* and *I. purpurea* found two *WD40* gene copies with high similarity to *AN11*. A frame shift mutation in one of these (*WDR1*) was found in a white flowered mutant and found to reduce expression of multiple genes encoding enzyme of the anthocyanin pathway in flowers (Morita *et al*., 2006).

The WD protein is located in the cytosol when expressed alone. By fractionating petal cell extracts, de Vetten *et al*. in 1997 showed that AN11 protein was found predominantly in the cytosolic fraction, not in the microsomal membrane fraction or in the nucleus. However, MBW activity was reported in nucleus (Sompornpailin e*t al*., 2002). It was believed that bHLH-WD40 dimerisation is necessary to transport WD40 into nucleus before forming the complete MBW complex in nucleus (Hichiri *et al*., 2011).

iv. The MYB-bHLH-WD40 (MBW) complex controls anthocyanin biosynthesis

Co-precipitation experiment by Zhoa *et al*. (2008) provided direct evidence for a MYB-bHLH-WD (MBW) complex that controls accumulation of proanthocyanidin

and trichome formation in Arabidopsis. Genetic evidence for the involvement of MBW complex that promoted anthocyanin pigmentation had been obtained from Petunia, involving AN2, AN1 and AN11 in the corolla, which mainly promoted expression of *DFRA, RT, AN9* and *CHSJ* but not *CHSA*, *CHI* or *AN3* expression (Huits *et al*., 1994; Kroon *et al*., 1994; Alfenito *et al*., 1998; Quattrocchio *et al*., 1993; de Vetten *et al*., 1997; Spelt *et al*., 2002). A similar requirement for control of different anthocyanin structural genes by MYB and bHLH genes was also found in *A. majus* flowers (Schwinn *et al*., 2006).

In *Antirrhinum*, expression of the the bHLH member (i.e., *Del* or *Mut*) appears to determine the tissue and developmental stage of anthocyanin gene expression (Schwinn *et al*., 2006), while the MYB partner might confer target gene specificity to the complex (reviewed by Hichiri *et al*., 2011).

In *Antirrhinum*, though knowledge of the control of pigment synthesis is still incomplete, the three anthocyanin regulatory genes *Del*, *El*, and *Ros* have been reported as having a major effect on pigmentation. The first two steps of anthocyanin and aurone biosynthesis, catalysed by CHS and CHI, show minimal regulation, but subsequent steps have an absolute requirement for the *Del* gene product in petal tubes and show quantitative regulation by *El* and *Ros*. Mutation of the *Del* gene reduces the level of transcripts of the genes encoding DFR (Almeida *et al*., 1989), F3H, ANS, and UF3GT (Martin *et al*., 1991) in the flower tube. The *Del* gene product was also proposed to acts as a repressor of *CHS* gene expression in the flower lobe mesophyll in which *del* mutant showed more *CHS* expression than in wild-type snapdragon (Jackson *et al*., 1992). However, this proposal is now considered unlikely because Jackson and colleagues themselves found that *Del* showed minimal regulation of *CHS* and *CHI*. Both *El* and *Ros* mutants have decreased expression of *F3H*, *DFR*, *ANS*, and *UF3GT* (Bartlett, 1989; cited by Martin and Gerats, 1993).

v. The MYB, MYB-bHLH regulation of flavonoid biosynthesis without WD-40 Repeat proteins

Although anthocyanin accumulation and some cell type specification mechanisms require a MBW complex, some proteins might act alone in regulating aspects of flavonoid biosynthesis. For example, AtMYB11, AtMYB12, and AtMYB111 appears to act alone in Arabidopsis in regulating *CHS*, *CHI*, *F3H*, *FLS* and *UFGT*, but not *DFR* expression (Stracke *et al*., 2007). Similarly, VvMYBF1 in grapevine regulates *VvFLS1* (*Flavonol Synthase 1*) expression without the need for a bHLH partner (Czemmel *et al*., 2009). In maize *ZmFLS1* expression is controlled by the anthocyanin promoting MYB–bHLH dimers, C1-PL1 and R-B, or by the phlobaphene promoting MYBP1 protein alone (Ferreyra *et al*., 2010). The available evidence therefore suggests that regulation of different parts of the flavonoid biosynthesis pathway may differ depending on plant species, and that this can involve either a MYB transcription factor acting alone or with a bHLH partner.

vi. Transcriptional regulation of the regulators

The members of the MBW complex can not only regulate flavonoid structural genes, but also their own expression or expression of other MBW genes in a complex loop. Examples of this phenomenon have been found in several species including Petunia, Arabidopsis and grape. The most commonly documented case involves up-regulation of the *MYB* or *bHLH* by the bHLH-MYB dimer. For example, *TT8* expression in Arabidopsis is promoted by TT2-TT8, PAP1-TT8 and PAP1-GL3 heterodimers (Tohge *et al*., 2005; Baudry, Chaboche and Lepiniec, 2006). In Petunia, the presence of AN2 and AN4 promotes *AN1* expression but has no effect on *JAF13* (Quattrocchio *et al*., 1998; Spelt *et al*., 2000). Hichri *et al*. (2010) also report that in grapevine, *VvMYC1* was up-regulated by VvMYC1-VvMYBPA1.

Evidence has been found to suggest that WD40s also participate in autoregulation. For example, in the Japanese morning glory (*I. nil*), *InbHLH2* expression was reduced in an *Inwdr1* mutant (Morita *et al*, 2006). Similarly in Petunia, DFR activity could be restored to *an11* mutant, lacking WD40 activity, by over-expressing *AN2*, which encodes its MYB partner. Although this was consistent with AN11 regulating

AN2, the study further found that *AN2* transcripts levels were not affected in an *an11* mutant, suggesting that AN11 might promote AN2 activity post-translationally (de Vetten *et al*., 1997).

1.3.4 Other loci that affect flower pigmentation in *Antirrhinum*

Two additional genes that are probably regulatory and known to affect flower pigmentation in *Antirrhinum* are *Eluta* (*El*) and *Sulfurea* (*Sulf*). *Eluta* (*El*) is a semi-dominant mutation that affects the pattern of colour in the flower – so that *A. majus* carrying a mutant *El* allele produces anthocyanin in the face of the flower and between dorsal petals. *El* is closely linked to *Ros*; raising the possibility that it is a *Ros* allele. The *Sulf* gene affects the distribution of yellow aurone pigments in the flower. In wild-type *A. majus*, aurones are restricted to the face of the flower, in *sulf* mutants they extend throughout the petals. *Sulf* is probably a regulatory gene because expression of the gene encoding auresidin synthase, which produces aurones from naringenin, is increased in *sulf* mutants. *sulf* mutations also decrease anthocyanin pigmentation, possibly because they allow substrate to be diverted from the anthocyanin pathway to the aurone pathway (Whibley, 2004).

1.4 Thesis aims

Antirrhinum species differ in flower colour and pattern, this variation is potentially important in their evolution (e.g, in reproductive isolation that lead to speciation). The aims of this thesis are to identify genes underlying natural flower colour variation, including the possibility of identifying novel genes and mechanism that effect anthocyanin biosynthesis and flower colour variation. The identified gene/s could reveal the molecular basis for natural variation which might cast further light on the regulation of anthocyanin biosynthesis.

Chapter 2

Materials and Methods

2.1 Plant material

2.1.1 Growing conditions

All plants were grown in the Institute of Molecular Plant Science glasshouses, at University of Edinburgh. Plants were grown in Levington F2 compost with 20% pf its volume of horticultural sand and maintained at 16°C-24°C with 16 hrs light (Hudson *et al*. 2009). Seed were germinated on the surface of compost and seedlings transplanted into 11cm x 11cm x 11 cm pot when the second pair of true leave were visible. If necessary, plants were potted on into larger pot and were trimmed occasionally to encourage production of more flowers.

2.1.2 Flower scoring

The *A. majus* x *A. charidemi* F_2 population, V1640, had been grown at John Innes Centre, Norwich and flowers of individual plants photographed by Dr Nicolas Langlade in 2002. These photographs were used to score corolla lobe and corolla tube pigmentation independently. Anthocyanin content of one petal lobe had also been estimated from the absorbance of acidified methanol extract (see below).

In near-isogenic lines (NIL), carrying regions of the *A. charidemi* genome in the genetic background of *A. majus*, three fresh flowers of each individual were scored to establish a standardized scale and then flowers from all individuals were photographed on a white background using same camera settings (taken during night time with a Nikon D60, with white balance =AUTO, ISO = 100, Exposure time =1/2"-1/20", F-number = 4-7). The camera was positioned 55 cm above the flowers. The same artificial lighting was used for all photos; the camera flash was not used. The flower lobe and tube tissues were then immediately separated and frozen in liquid nitrogen, and stored in -80 ^{o}C for later RNA extraction.

Flower used for UV-visible spectrophotometry were photographed separately in natural daylight (Nikon D60, white balance =AUTO, ISO = 100, Exposure time = AUTO, F-number = AUTO) to record the intensity and pattern of anthocyanin pigmentation in the petal lobe of each individual.

The intensity of pigmentation in petal lobes was scored on a scale from 1 to 10 and flower tube pigmentation on a scale of 1 to 6. Scoring charts were generated from images of representative flowers in each category to ensure uniformity (Figure 3).

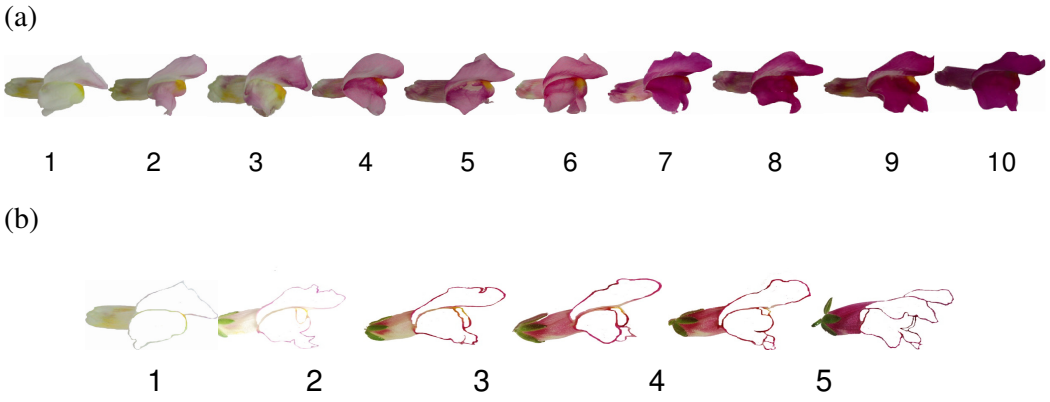


Figure 3. Flower scoring standard (a) The array of petal lobe colour scores from 1 (palest) to 10 (darkest). (b) The array of petal tube colour scoring from 1(palest) to 6 (darkest).

2.2 DNA methods

2.2.1 DNA extraction

Young vegetative shoot tissue (about 0.1 g if available) was collected in a 1.5 ml Eppendorf tube with two 4 mm steel balls, cooled in ice-box and then frozen in liquid nitrogen as soon as possible. Tissue was then used for extraction of stored at -80^{o}C until required. Different method of DNA extraction were used, according to the quantity and quality of DNA required.

i. CTAB miniprep (Doyle and Doyle, 1990)

The frozen leaf tissue was ground by shaking twice in a Retsch Mixer Mill MM300 (RETSCH Ltd. *http://www.retsch.com*) at 30 Hz for 1 minute each. Extraction buffer

(0.5 ml of 100 mM Tris pH 8.0, 1.4M NaCl, 20 mM EDTA, 2% CTAB, 0.2% β-mercatoethanol) was added and the mixture homogenized by shaking again in the mixer mill at the same speed for another minute. The tube was then incubated at 65^{o}C for 30 minutes after which 0.5 ml of chloroform ($CHCl_3$) was added. The contents of the tube were mixed by vortexing for 20 sec. then centrifuged at 13,000 x *g* for 5 minutes. The aqueous supernatant was transfer to a new Eppendorf tube and 334 µl of isopropanol added. The contents of the tube were mixed by vortex and again centrifuged at 13,000 x *g* for 10 minutes to pellet nucleic acids. The supernatant was removed and the pellet dissolved in 50 µl of TE (10 mM Tris, pH 7.4, 0.1 M EDTA) containing 10 µg/ml of RNase. The tube was incubated for an hour at 37^{o}C to digest unwanted RNA. DNA was then precipitated from the solution by addition of 2.5 volumes of absolute ethanol and 0.1 volume of 3 M sodium acetate (pH 7.0) and recovered by centrifugation at the same speed for 5 minutes. The DNA pellet was washed with 70% ethanol, air-dried, and dissolved in 50 µl of sterile deionised water. DNA yield and quality was measure by with a ND-1000 Nanodrop Spectrophotometer (Nanodrop Technologies, Inc., *http://www.nanodrop.com*), with the assumption that 1 unit of absorbance at 260 nm corresponds to 50 µg/ml of DNA. DNA solutions were stored at -20 ^{o}C.

ii. Edwards miniprep (Edwards *et al.*, 1991)

Frozen leaf tissue was ground as previously described. Extraction buffer (0.4 ml of 200 mM Tris pH 7.5, 250 mM NaCl, 25 mM EDTA, 0.5% SDS) was added, mixed with the sample and then left at room temperature for up to 1 hour or until the supernatant appeared clear. The mixture was vortexed again, left at room temperature for another 2 minutes then centrifuged at 13,000 x *g* for 2 minutes. The supernatant was removed to new tube and an equal volume of isopropanol added to the supernatant to precipitate DNA. The contents of the tube were vortexed, incubated at room temperature for 2 minutes and then centrifuged at 13,000 x *g* for another 2 minutes. The supernatant was discarded and the DNA pellet air-dried for 15-60 minutes. The dried DNA pellet was dissolved in 10 mM Tris (pH 8.0). DNA yields and quality were estimated using the Nanodrop as previously described.

iii. Phenol:chloroform miniprep (adapted from Sambrook, Fritsch and Maniatis, 1989)

Frozen leaf material was ground as previously described and mixed with 0.5 ml of extraction buffer (100 mM Tris pH 8.0, 100 mM NaCl, 50mM EDTA, 1% SDS) . The mixture was incubated at 65^{o}C for at least 10 minutes then allowed to cool briefly. An equal volume of a 1:1 (v/v) mixture of phenol and chloroform was added and the mixture vortexed thoroughly. The mixture was then centrifuged for 1 minute at 13,000 x *g* and the aqueous phase transferred to a fresh Eppendorf tube. The phenol:chloroform extraction step was then repeated. Nucleic acid was precipitated from the aqueous phase by addition of 0.1 volume of 3M sodium acetate, pH 5.5 and 0.7 volume of isopropanol. Following thorough mixing, nucleic acids were recovered by centrifugation at 13,000 x *g*. The supernatant was discarded and the pellet washed with 0.5 ml 70% ethanol and then air-dried. It was then left overnight to dissolve in 50 µl TE or water at 4^{o}C. DNA yield and quality was estimated by Nanodrop.

2.2.2 Restriction endonuclease digestion of DNA

Sequence information was used to develop cleaved amplified polymorphic sequence (CAPS) markers for genotyping. Digestion reaction was normally carried out by adding 4 µl of a PCR product to a reaction mix containing 1-2 units of restriction enzyme (supplied by New England Biolabs; NEB), 1 µl of the 10x buffer and 0.1 µl BSA solution supplied with the enzyme and sterile deionised water to make 10 µl total volume. Reactions were incubated as instructed by enzyme supplier. The products then were separated by agarose gel electrophoresis, as described below.

2.2.3 Agarose gel electrophoresis of DNA

Digested DNA, or PCR products were separated by electrophoresis in gels containing 1.5-2.5% (w/v) agarose in 0.5x TBE buffer (0.089 M Tris base, 0.089 M boric acid and 0.022 M EDTA). DNA samples were prepared by adding 0.1 volume of loading buffer (0.25% bromophenol blue, 0.25% xylene cyanol and 30% glycerol in water). Typically 4 µl of each sample was loaded in a 3 mm deep meniscus gel.

To allow estimatation of the sizes of fragments obtained, about 125 ng of DNA size ladded was also loaded into one lane of each gel. Gels were run at 5-8 V/cm, for 45-90 minutes, depending on size of the fragments being analysed and the resolution required. Ethidium bromide added during gel preparation to a concentration of 0.5 μg/ml allowed DNA to be visualized when illuminated by UV light using Geneflash (Syngene) UV transilluminator and photographed by the same machine.

2.2.4 Polymerase chain reaction (PCR) techniques

i. Primer design

New primers were designed by the Primer3 software (available at http://www.broad.mit.edu/cgi-bin/primer/primer3_www.cgi; Rozen and Skaletsky, 2000).

ii. Standard PCR protocol

Approximately 15-25 ng of genomic DNA was used as a template for standard PCR. General mixtures for 10 μl reactions were set up with 0.5 unit of of *Taq* DNA polymerase (5 units/μl; New England Biolabs), 4 pM each of forward and reverse primers , 4 nM of each dNTP, 1x PCR buffer (supplied with the enzyme, 10 mM Tris-HCl, 50 mM KCl, pH 8.3 and 1.5 mM $MgCl_2$, unless otherwise stated). Alternatively 1 μl of homemade *Taq* polymerase was used and the buffer replaced with 1x yellow PCR buffer (50mM Tris-HCl, pH 8.3, 500 μg/ml BSA, 0.5% Ficoll, 1% sucrose, 30mM KCl, 1mM $MgCl_2$ and 1mM tartrazine chloride). In both cases, sterile deionised water was added to make up the volume to 10 μl.

PCR was carried out with an initial 94°C denaturation step for 2 minutes; followed by between 25 and 40 cycles of a 94°C denaturation for 1 minute, annealing for 2 minutes at between 50°C and 65°C (depending on the primers and template) and a 72°C extension for between 45 seconds and 2 minutes 30 seconds (assuming a synthesis rate of 1 kb/min). Reactions were completed with a final extension at 72°C for 10 minutes. All thermocycling was carried out in a PTC-200 Peltier Thermal Cycler (MJ Research). Samples were then stored at 4°C or -20°C. The sequences of

gene-specific primers used in the course of this study and their specific PCR conditions are presented in Tables 1 and 2.

2.2.5 Cloning into plasmid vectors and colony PCR

PCR products were cloned by ligation in to the pJET1.2/blunt vector (supplied in a kit by Fermentas). This vector has a blunt-end insertion site for PCR products and also several flanking restriction sites suitable for subcloning PCR products. The vector carries ampicillin and kanamycin resistance genes, allowing antibiotic selection of bacteria transformed with this plasmid. In addition, the multiple cloning region is located within the coding region of the *β*-galactosidase *α*-peptide allows clones with an insert to be detected as white colonies, compared to clones lacking an insert that appear blue on medium containing X-gal and IPTG. Cloning was performed by removing the 5' overhanging A base from PCR products using the blunting enzyme supplied with the kit at 70^{o}C for 5 minutes. The PCR product was then ligated into pJET1.2/Blunt cloning vector at room temperature for 5 minutes using the *T4* DNA ligase supplied. The ligated vector mixture was used directly in heat-shock transformation of *E. coli*.

i. Heat-shock transformation

Competent cells of the DH5*α* strain *E. coli* were kindly provided by Andrew Waters. The competent cell suspension, which had been stored at -80°C, was melted on ice. Ligtation mixture (5 µl) was added to 100 µl of competent cells and mixed gently. The cells were incubated on ice for 20-30 minutes then heat-shocked in a waterbath at 42^{o}C for 90 seconds and returned to ice for 2 minutes. LB-both (900 µl) was added to each tube and the cells were incubated at 37^{o}C for 1 hour with gentle agitation. Cells were then pelleted by centrifugation at low speed (~10 x *g*). All except about 100 µl of the medium was discarded and the bacterial pellet was gently re-supended in the remainder. The suspension was then spread onto LB-agar plates [10 g/l Bactotryptone; 5 g/l yeast extract; 5 g/l NaCl; 1 g/l D-glucose; 12 g/l agar] containing 100 µg/ml ampicillin (added from a 100 mg/ml stock in deionised water) without X-gal and IPTG. Plates were incubated upside down at 37^{o}C overnight.

Primer number	Primer name	Anneals to	Primer sequences	Optimum annealing temp./extension time	Function (analyse by)	Amplification size	Note
209	NIVF	*Nivea*	TAGCACAACCACTTCCGTCACC	55°C/1 m 30 s	CAPS marker (*Mse* I) [2]	Many unexpected bands	Differ around 0.3-0.4 kb band size after digested
210	NIVR	*Nivea*	ACACGAGCACCGGCATTGTTTT				
1013	NIV1inF	*Nivea*	ACAGTTTTGGCGATTGGAAC	60°C/1m		~1.1kb	Partial 1st exon, 1st intron, partial 2nd exon. No sequence
1014	NIV1InR	*Nivea*	CACATGGCAGGGTTCTCTTT				
1015	NIV2InF	*Nivea*	CAAAGAAGCTGCCCAAAAAG	60°C/1m		~500bp	Partial 2nd exon, 2nd intron, partial 3rd exon. No SNP
1016	NIV2InR	*Nivea*	AACAAACGACCAACACACGA				
1032	CHI1InF	*CHI*	GGCTGAGATCACCCAAATTC	50°C/30 s		~300bp	Partial 1st exon, 1st intron, partial 2nd exon. No sequence
1033	CHI1InR	*CHI*	TCTTGCCCTTCCACTTAACG				
720	G425	*Incolorata*	ATGGCTCCAATGCCAACATCAC	60 °C/30 s		~350 bp	5'UTR and partial 1st exon
982	1R	*Incolorata*	CTGAAGATGGCTAGAGACGATG				
721	NOR	*Incolorata*	GACGAAGACGAGAGGCCTAAGG	50 °C/1 m		~700 bp	Partial 1st exon and 1st intron
722	R427	*Incolorata*	GTACCCGGATCCGTATGTCGCTTTA				
776	INCF426	*Incolorata*	AGCGACATACGGATCAGGGTAC	61 °C/1m 30 s	CAPS marker[3] (size difference)	~ 1.32 kb (*A. molle*) ~1.35 kb (*A. majus*) ~ 1.41 kb (*A .charidemi*)	Partial 2nd exon, 2nd intron, partial 3rd exon
777	INCREX3	*Incolorata*	GTTCACAACTGCTTGGTGGTC				
981	G1058	*Incolorata*	GACCACCAAGCAGTTGTGAACTC	62 °C/30 s		~500 bp	Partial last exon and 3' downstream
984	G340	*Incolorata*	GGAGAGGATGACATATGTCCAAA				

1= CAPS markers between *A. majus* and *A. molle* alleles
2 =CAPS markers between *A. majus* and *A. charidemi* alleles
3 =CAPS markers between *A. majus*, *A. charidemi* and *A. molle* alleles

Table 1. Primers used to amplify CAPs genotyping markers or to identify allelic variation by sequencing (to be continued).

Primer number	Primer name	Anneals to	Primer sequences	Optimum annealing temp./extension time	Function (analyse by)	Amplification size	Note
1021	Eosina3InF	*Eosina*	TTTGCTGCAGGAACTGACAC	55 ºC/30 s		~300bp	Partial 3rd exon, 3rd intron, partial 4th exon. No SNP
1022	Eosina3InR	*Eosina*	TCGAACCCTTTGGAATCAAG				
780	PLDA_FW_12	*Pallida*	TACCGGTGGTGAGTTTCTCC	58 ºC/30 s		~450 bp	Partial last exon
781	PLDA_RVS_390	*Pallida*	TGCTTAACCCCTGAAGTTGG				
985	Intron1_PallidaR	*Pallida*	CGTGTCTGCTTTTGGCAGT	50 ºC/30 s		~400 bp	1st intron
986	Intron1_PallidaF	*Pallida*	CCACCGTATGTGTCACAGGA				
987	IntronLast_ PallidaR	*Pallida*	AACCCATTCCGATCATTTTCT	52 ºC/30 s	CAPS marker (*Stu* I)[2]	~500 bp	Partial 5th exon, Last intron, Partial 6th exon
988	IntronLast_ PallidaF	*Pallida*	GAGAACTGGCCCGAATATCA				
199	CANDIF	*candica*	TACATCCGACCCGAAGAAGAGTTA	55ºC/1 m 30 s	CAPS marker (*Taq* I) [2]	Many unexpected bands	Differ around 0.7-0.9 kb band size after digested
200	CANDIR	*candica*	AGGGCATTTTGGGTAGTAGTAGTTGAT				
1019	Candi?RTF	*Candica*	CTAACAACGCGAGTGGACAA	55ºC/30 s		~300 bp	Designed from *TDS4* gene in *Arabidopsis* No SNP
1020	Candi?RTR	*Candica*	ACCGACAGAGAGAGCCTTGA				
1023	AJ796511F	*PAT1*	TTATTTTCACTGGCTATCATATTTCC	55ºC/1 m		~1.1kb	No SNP
1024	AJ796511R	*PAT1*	ATCCCAATCAAAAGCGACTG				
1025	AJ795263F	*PAT2*	GTGGTTGGCGTTTCAAGACT	55ºC/1 m		~900 bp	No restriction site available
1026	AJ795263R	*PAT2*	CTAAAACAACTGCCGTGCAA				

1= CAPS markers between *A. majus* and *A. molle* alleles
2 =CAPS markers between *A. majus* and *A. charidemi* alleles
3 =CAPS markers between *A. majus*, *A. charidemi* and *A. molle* alleles

Table 1 (continued). Primers used to amplify CAPs genotyping markers or to identify allelic variation by sequencing (to be continued).

Primer number	Primer name	Anneals to	Primer sequences	Optimum annealing temp./extension time	Function (analyse by)	Amplification size	Note
1029	AJ794241F	*PAT4*	GCATTGGAGACACTGTGTGG	55°C/30 s		~250bp	No SNP
1030	AJ794241R	*PAT4*	AGAACCTCTGCAGTGGGAAA				
1043	AmFLS1F	*Flavonol synthase*	AAGGATGGGTGGATCATTTG	50 °C/30 s		~1 kb	Partial 2nd exon, 2nd intron, partial 3rd exon. No avail. restriction site
1044	AmFLS1R	*Flavonol synthase*	CCTTCAACTCATGCCCTTTT				
1017	DELRTF	*Delila*	AATTATGGCGCAACAAGGAC	55°C/30 s		~500bp	No SNP
1018	DELRTR	*Delila*	CTTGCTGATGCAACCTTCAA				
994	MutProF	*Mutabilis*	TTCGTCGCCACCTAATTTTC	50 °C/1 m		~900bp	5'UTR, partial 1st exon
995	MutProR	*Mutabilis*	TGATCAGTGGCATTTCCGTA				
996	Mut1InF	*Mutabilis*	AAGAGTCGGCTTCAGTGCAT	50°C/30 s	CAPS marker (*Bcl* I)[2]	~500 bp	Partial 1st exon, 1st intron, partial 2nd exon
997	Mut1InR	*Mutabilis*	TTCCGTGCTCACTTCTGTTG				
998	Mut2ExInF	*Mutabilis*	TCTAGTGTGGCGTGATGGAT	50 °C/1 m		~1.2 kb	Partail 2nd exon , 2nd intron, partial 3rd exon
999	Mut2ExInR	*Mutabilis*	ATCCATACATGCTGCCGTTT				
1000	WD40_1cDNAF	*WD40*	GACATTTCAACCCACGATCC	52°C/30 s	CAPS marker (*Msp* I) [2]	~450bp	Coding region of WD40
1001	WD40_1cDNAR	*WD40*	AAACGAAGTCAACGGAGCAC				
1002	WD40_2cDNAF	*WD40*	GTGCTCCGTTGACTTCGTTT	52°C/ 45 s		~700bp	Coding region of WD40
1003	WD40_2cDNAR	*WD40*	TTTCAGCATCTGCATCTTGC				

1= CAPS markers between *A. majus* and *A. molle* alleles
2 =CAPS markers between *A. majus* and *A. charidemi* alleles
3 =CAPS markers between *A. majus*, *A. charidemi* and *A. molle* alleles

Table 1 (continued). Primers used to amplify CAPs genotyping markers or to identify allelic variation by sequencing (to be continued).

Primer number	Primer name	Anneals to	Primer sequences	Optimum annealing temp./extension time	Function (analyse by)	Amplification size	Note
100	ROSEAF	*Rosea1 (Reverse)*	CTCTGTGTGGAACTTGAT	52 °C/1 m	CAPS markers[3] (size differences)	~ 300 bp (*A. molle*) ~500 bpp (*A. majus*) ~ 1 kb (*A. charidemi*)	5'UTR and partial 1st exon
101	ROSEAR	*Rosea1 (Forward)*	CTCCCTACTGTATCTTGG				
474	ROSEAF1	*Rosea1*	GAACACAAAAACCATTAAGCAGACTAA	57 °C/30 s	CAPS marker (*Hpa* II) [1]	~320 bp	Last exon
481	ROSEAR	*Rosea1*	ATTTCCAATTTGTTGGGCCTCG				
1034	VE2ExonF1	*venosa*	TCGAGAAGAAGTCCGGAGAA	55°C/30 s	CAPS marker (*Dde* I) [2]	~300bp	2nd exon
1006	Ve2Ex_LastExR	*Venosa*	ACGTAGTCCGCAAAATCGAG				

1= CAPS markers between *A. majus* and *A. molle* alleles
2 =CAPS markers between *A. majus* and *A. charidemi* alleles
3 =CAPS markers between *A. majus*, *A. charidemi* and *A. molle* alleles

Table 1 (continued). Primers used to amplify CAPs markers and in search of allelic variation in genes involved in anthocyanin biosynthesis.

Primer number	Primer name	Anneals to (orientation)	Primer sequences	Optimum annealing temp./extension time	Function (analyse by)	Amplification size	Note
5	MIXL1R3	*mixta-like1* (Reverse)	AAAGGCATGAAGTTGGATAATAGC	60 °C/1 m	CAPS marker (*Hpa* II)[2]	~700 bp	
9	MIXL1 F3	*mixta-like1* (Forward)	CCAGAAGAAGACCAAAAACTATTGT				
72	FAP2F	*FAP2*	TCGCCTTGTTGTCTCAAACGAT	45 °C/1 m 30 s	CAPS marker (*Hae* III)[2]	Many unexpected bands	Differ around 1.2-1.3 kb band size after digested
73	FAP2R	*FAP2*	GAACGCCCACTGGTTCTCTCTT				
191	AMANTF1	*Amaintegumenta*	GATACATTTGGGCAGAGAAC	55 °C/1 m 30 s	CAPS marker (*Rsa* I)[2]	Many unexpected bands	Differ around 0.7 -0.8 kb band size after digested
192	AMANTR1	*Amaintegumenta*	GCAGCTTTCTCTTCCATATC				
233	5_08_o20F		TTACCTGTGCCTCTCCTTC	55 °C/1 m 30 s	CAPS marker (*Mse* I)[2]	Many unexpected bands	Differ around 0.7-0.8 kb band size after digested
234	5_08_o20R		AAATAATGACCAAAAACTGCTT				

Table 2. Primers used to amplify markers in LG7

Colonies were picked from the plate with a sterile yellow tip and screened for the presence of recombinant plasmids by colony PCR.

ii. Colony PCR

Each single colony was dispersed in 15 µl of sterile deionized water in a PCR tube. The tubes were incubated at 95^{o}C for 5 minutes then vortexed and centrifuged at 20,000 x *g*. The supernatants were used as templates for PCR (3 µl of supernatant in a 25 µl PCR).

Colony PCR mixtures of 25 µl were set up with 3 µl of heated colony mixture as template, 10 pM each of pJET1.2 forward and reverse primers (pJET 1.2 Forward primer; 5'-CGACTCACTATAGGGAGAGCGGC-3', pJET 1.2 Reverse primer; 5'-AAGAACATCGATTTTCCATGGCAG-3'), with other components as described previously.

PCR was carried out with an initial 94^{o}C denaturation step for 2 minutes; followed by 35 cycles of 94^{o}C denaturation for 30 seconds, 60^{o}C annealing for 30 seconds and 72^{o}C extension for 1 minute 30 seconds and then final extension step at 72^{o}C for 10 minutes.

2.2.6 DNA sequencing

Sequencing reactions were carried out by the University of Edinburgh sequencing service (The GenePool). Samples were prepared by adding 0.6 µl of a PCR product and 1 µl of forward or reverse primer (10 µM) to 4.4 µl of water.

2.2.7 Amplified Fragment Length Polymorphism (AFLP)

Amplified fragment length polymorphisms were used to identify molecular markers linked to flower colour loci.

The AFLP method was based on the approach described in Vos *et al.* (1995), with many modifications. It consisted of several steps, as described below;

i. Template generation step

a) Digestion of genomic DNA

Approximately 125 µg of genomic DNA was digested with 2.5 units each of *Mse* I and *Pst* I restriction enzymes in a total reaction volume of 10 µl. The reaction also contained 1x Restriction-Ligation (RL) buffer (5x RL is 50 mM Tris-acetate pH7.5, 50 mM Mg acetate, 250 mM K acetate; 25 mM DTT (dithiothreitol), 250 ng/µl BSA (bovine serum albumin) and was made by supplementing 10x *OnePhorAll* buffer with BSA and DTT). The digestion reaction mixture was incubated at 37°C for 2 hours.

b) Ligation of Genomic DNA and adapter

- *Preparation of adapters*

The adapters were prepared by heating equal molar amounts of *Mse* I adapters oligonucleotides (50 µl of 100 mM solutions of forward and reverse oligos; Table 3a) with 1/10th volume of 10x *OnePhorAll* buffer (100mM Tris.HAc pH 7.5, 100mM MgAc, 500mM KAc) at 65^{o}C for 10 minutes, 37^{o}C for 10 minutes and then 25^{o}C for another 10 minutes. Oligonucleotides forming the *Pst* I adapter (Table 3a) were annealed in the same way and then diluted 1:10 with sterile deionised water. Adapters were freshly annealed for each AFLP experiment.

- *Ligation of adapters*

The adapters were ligated to the digested genomic DNA by adding a ligation mix (2.5 µl) to the 10 µl digestion reaction. The ligation mix contained 0.5 µl of fresh 5x RL buffer, 0.25 µl of prepared *Mse* I adapter (50 µM), 0.25 µl of prepared *Pst* I adapter (5 µM), 0.25 µl of rATP (10 mM) and 0.0625 µl of T4 DNA ligase (New England Biolabs; 400 U/µl) with water to make up the volume to 2.5 µl. The reactions were thoroughly mixed then spun down briefly before incubation at 37^{o}C for 3 hours or at 16°C overnight.

ii. Pre-selective PCR amplification

Pre-selective amplification was carried out in a 20 µl reaction volume containing 5 µl of adapter-ligated template, 0.1 µl of *Taq* DNA polymerase (5 units/µl; New England

Biolabs), 1x PCR buffer (Mg-free, supplied with the enzyme), 1.6 µl of 25 mM $MgCl_2$, 0.4 µl of 10 mM each dNTPs and 0.6 µl of each pre-selective amplification primer, M00 and P00 (10 µM each; Table 6). Water was then added to make 20 µl total volume. PCR conditions consisted of a 72°C incubation for 2 minutes then 20 cycles of denaturing at 94°C for 20 seconds, annealing at 56°C for 30 seconds and extension at 72°C for 2 minutes. This was followed by a final extension step at 72°C for 30 minutes. The success of digestion, ligation and pre-selective amplification was checked by separating 3 µl of the PCR in a 1.5% agarose gel. Numerous PCR products between 100 bp and 800 bp indicated that the reaction was successful.

iii. Selective amplification

To select a subset of fragment, selective PCR was carried out with primers that were complementary to the P00 primer but had an additional two 3' nucleotides beyond the *Pst* I site (Table 3a). These primers were fluorescently labelled at their 5' ends with one of the dyes compatible with the ABI (Applied Biosystems) filter set G5 (6'FAM, PET, VIC and NED). Similarly the *Mse* I primers were complementary to M00 but carried additional selective 3' bases (AGA, ACA, CAC, AGG, ACG, AGC or CAT; details in Table 6a). The different combinations of primers are given in Table 3b.

Selective amplification was carried out in 10 µl reactions containing 1 µl of pre-selective amplification reaction, 0.25 µl of each selective primer (10 µM), and other reagents as detailed previously. The PCR reaction was carried out using a touch-down programme that consisted of one step of initial denaturation for 2 minutes at 94°C, follow by 10 cycles of 94°C denaturation for 20 seconds, annealing for 30 seconds initially at 66°C but decreasing by 1°C each cycle, and extension at 72°C for 2 minutes. The remaining 20 cycles consisted of denaturing at 94°C for 20 second, annealing at 56°C for 30 and extension at 72°C for 2 minutes. A final extension step 72°C for 30 minutes ended the programme.

iv. Preparation of samples for separation in the ABI3730 sequencer

Four selective PCRs using *Pst* I selective primers labelled with different dyes were mixed, using 1 µl of each PCR reaction, and spun down. 1 µl of of the mixture added to 9 µl of Hi-Di Formamide mix (ABI) with 1 µl/ ml of added LIZ-500 size standard (ABI).

Samples were separated in an ABI3070 sequencer at The GenePool. Trace files were returned in .fsa format (ABI) and converted to virtual gels in several steps. The first involved conversion of the file format using the BF30 FSA Converter (University of Madison, Wisconsin), calibration with the internal LIZ-500 size standard, using GeneScan software (ABI) and loading into the 5-colour Genographer programme (Bentham, 2001; available at http://hordeum.oscs.montana.edu/genographer). Genescan converted ABI trace files into virtual gel images and enabled band scoring and accurate determination of band sizes.

(a)

Primer function	Primer name : Sequence
Mse adapter strands	MseI AdF: GACGATGAGTCCTGAG
	MseI AdR: TACTCAGGACTCAT
Pst adapter strands	PstI AdF: CTCGTAGACTGCGTACATGCA
	PstI AdR: TGTACGCAGTCTAC
Pre-selective primer	M00: GATGAGTCCTGAGTAA (complementary to Mse adapter) P00: GACTGCGTACATGCAG (complemetary to Pst adapter)
Pst selective primers	P11: GACTGCGTACATGCAGAA P12: GACTGCGTACATGCAGAC P14: GACTGCGTACATGCAGAT
Mse selective primers	Mse-ACA: GATGAGTCCTGAGTAAACA Mse-ACG: GATGAGTCCTGAGTAAACG Mse-AGA: GATGAGTCCTGAGTAAAGA Mse-AGC: GATGAGTCCTGAGTAAAGC Mse-AGG: GATGAGTCCTGAGTAAAGG Mse-CAC: GATGAGTCCTGAGTAACAC Mse-CAT: GATGAGTCCTGAGTAACAT

(b)

ABI3730 channel	Pst primer-fluorescent label	Mse primer
blue	P11-6FAM	Mse-AGA/ Mse-CAC
red	P11-PET	Mse-CAT/ Mse-AGG
Green	P12-VIC	Mse-ACA/ Mse-ACG
yellow	P14-NED	Mse-ACA/ Mse-AGC

Table 3. Primers used in AFLP (a) Primers sequences, (b) combination of primers in selective-amplification used in this project.

2.2.8 Transposon Display (adapted from Souer *et al*., 1996)

Genomic DNA of F_2 hybrids (*A. majus* x *A. molle*) that showed similar ranges of anthocyanin content were pooled. DNA (200 ng) was digested with *Pst* I or *Mse* I in 400 µl at 37°C for 2 hours 30 minutes, as described previously. Digests were then incubated at 65°C for 20 minutes to inactivate the enzymes. Ligation was carried out in 40 µl contained 10 µl of the digested DNA mixture, 4 µl of 10x ligase buffer (NEB), 0.1 µl of T4 DNA ligase (400 U/µl) and 25.9 µl deionized autoclaved water. The ligation mixture was incubated overnight at 4°C then used as the template for PCR. PCR reactions were prepared as describe for Standard PCR, using *Tam2* primers (TAM2LA [forward]; 5'-CGTGCAGTTGGTGGCTGGTA-3', TAM2LT [reverse]; 5'-ACCCATGCACTCTTGGGACA-3'). PCR was carried out with an initial 94°C denaturation step for 10 minutes followed by 45 cycles of 94°C denaturation for 20 seconds, 56°C annealing for 1 minute and 68°C extension for 2 minutes 30 seconds. A final extension at 68°C for 10 minutes completed the PCR. Products were analysed by separation in 2% agarose gels in 0.5x TBE at 8 V/cm for 15 hours.

2.3 RNA methods

2.3.1 Total RNA extraction

Frozen petals were ground as described for DNA extraction. 1 ml of TRIzol® Reagent (Life Technologies) was added immediately and mixed thoroughly with the samples. Tubes were then incubated in room temperature for 5 minutes and centrifuged for 5 minutes at low speed (6,000 x *g*) to precipitate debris. The supernatant was transferred to a fresh RNase-free Eppendorf tube and spun at 12,000 g for 10 minutes at 4 °C. The supernatant was transferred again to a fresh RNase-free Eppendorf tube and 0.2 volume of chloroform added. The mixture was vortexed and spun at 20,000 x *g* for 10 minutes at 4°C. The aqueous phase was transferred to a fresh RNase-free Eppendorf tube and RNA precipitated by addition of 0.5 ml of isopropanol. The RNA was recovered by centrifugation at 20,000 x *g* for 15 minutes at 4°C. The supernatant was discarded and the RNA pellet was dissolved in 100 µl

warm (65°C) DEPC-treated water. RNA concentration was measured by Nanodrop. RNA quality was determined by separation in agarose gels (1.5% agarose, 8 V/cm, 30 minutes) using equipment that had been made free of RNase by washing in 1 M NaOH solution.

2.3.2 First strand cDNA synthesis and RT-PCR

First strand cDNA synthesis and RT-PCR were conduct using a two-step dART cDNA Synthesis/RT PCR kit (*EUR*x). Synthesis of single-stranded cDNA was carried out by first making a 13 µl mix containing 1µl of oligodT_{20} (50 µM), 2.5 µg of total RNA, 4 µl of dNTP mix (5 mM each) and RNase-free water to make up the volume. The mix was heated at 65°C for 5 minutes, to denature any secondary structure, and then chilled immediately on ice for another 5 minutes. Then 7 µl of a cDNA master mix was added, this contained 4µl of 5x cDNA buffer, 1µl of RNase-inhibitor (12.5 U), 1µl of DTT (100mM) and 1µl of dART Reverse Transcriptase. The reaction was incubated at 50°C for 2 hours then heated at 85°C to terminate reverse transcriptase activity. The mixture was used as the cDNA template for the second step (PCR). PCR contained 5 µl of cDNA template, 1µl of *OptiTaq* DNA polymerase (2.5 U/µl), 5 µl 10x *Pol* buffer C with $MgCl_2$, 2 µl of dNTPs (5 mM each) and 1 µl of each forward and reverse primer (10 µM; see Table 4) with water to make 50 µl total volume. PCR was carried out with an initial 94°C denaturation step for 3 minutes; followed by between 30 and 35 cycles of 94°C denaturation for 30 seconds depending on the primers, a 30 second annealing step at between 50°C and 55°C depending on the primers and 72°C extension for 1 minute 30 seconds. One final extension cycle at 72°C for 10 minutes completed the PCR.

Primer number	Primer name	Anneals to (orientation)	Primer sequences	Optimum annealing temp./extension time/cycles	Amplification size	Note
776	INCF426	*Incolorata (Forward)*	AGCGACATACGGATCAGGGTAC	55 °C/30 s	~200 bp	Also used as CAPS marker
777	INCREX3	*Incolorata (Reverse)*	GTTCACAACTGCTTGGTGGTC			
724	5'ROS1	*Rosea1*	AAAGGATCCATGGAAAAGAATTGTCGTGGAGTG	50 °C/30 s	~700 bp	Not in Database catalog
725	3'ROS1END	*Rosea1*	AAAGGATCCTTAATTTCCAATTTGTTGGG			
726	5'ROS1	*Rosea2*	AAAGGATCCATGCAAAAGAATCCTCGGGGAGTA	50 °C/30 s	~700 bp	Not in database catalog
727	3'ROS1END	*Rosea2*	AAAGGATCCTTACCATTCAGACATTTTTCCGG			
987	IntronLast_ PallidaR	*Pallida*	AACCCATTCCGATCATTTTCT	50 °C/30 s	~100 bp	Also used as CAPS marker
988	IntronLast_ PallidaF	*Pallida*	GAGAACTGGCCCGAATATCA			
996	Mut1InF	*Mutabilis*	AAGAGTCGGCTTCAGTGCAT	50°C/30 s	~100 bp	Also used as CAPS marker
997	Mut1InR	*Mutabilis*	TTCCGTGCTCACTTCTGTTG			
1000	WD40_1cDNAF	*WD40*	GACATTTCAACCCACGATCC	50°C/30 s	~450bp	*WD40* contain no intron.
1001	WD40_1cDNAR	*WD40*	AAACGAAGTCAACGGAGCAC			
1034	VE2ExonF1	*venosa*	TCGAGAAGAAGTCCGGAGAA	50°C/30 s	~300bp	2nd exon /also used as CAPS marker
1006	Ve2Ex_LastExR	*Venosa*	ACGTAGTCCGCAAAATCGAG			
1017	DELRTF	*Delila*	AATTATGGCGCAACAAGGAC	5550°C/30 s	~300bp	
1018	DELRTR	*Delila*	CTTGCTGATGCAACCTTCAA			
1015	NIV2InF	*Nivea*	CAAAGAAGCTGCCCAAAAAG	55°C/30 s	~200bp	
1016	NIV2InR	*Nivea*	AACAAACGACCAACACACGA			

Table 4. RT-PCR primers for genes involved in anthocyanin biosynthesis (to be continue).

Primer number	Primer name	Anneals to (orientation)	Primer sequences	Optimum annealing temp./extension time/cycles	Amplification size	Note
1019	Candi?RTF	*Candica*	CTAACAACGCGAGTGGACAA	55°C/30 s	~200 bp	Designed from *TDS4* gene in *Arabidopsis*
1020	Candi?RTR	*Candica*	ACCGACAGAGAGAGCCTTGA			
1021	Eosina3InF	*Eosina*	TTTGCTGCAGGAACTGACAC		~300bp	
1022	Eosina3InR	*Eosina*	TCGAACCCTTTGGAATCAAG			
1023	AJ796511F	*PAT1*	TTATTTTCACTGGCTATCATATTTCC	55°C/30 s	~600bp	Design from AJ796511 homolog to tomato PAT1
1024	AJ796511R	*PAT1*	ATCCCAATCAAAAGCGACTG			
1032	CHI1InF	*CHI*	GGCTGAGATCACCCAAATTC	50°C/30 s	~200bp	
1033	CHI1InR	*CHI*	TCTTGCCCTTCCACTTAACG			
1043	AmFLS1F	*Flavonol synthase*	AAGGATGGGTGGATCATTTG	50 °C/30 s	~100bp	
1044	AmFLS1R	*Flavonol synthase*	CCTTCAACTCATGCCCTTTT			
321	AmA2R	*Nivea*	CATAAGGTAGTCAGTGAGGTC	55°C/30 s	~500bp	RT-PCR control
322	AmA2F1	*Nivea*	GAGATAATGGGAACTGGAATGG			
1041	UBI.F	*Ubiquitin*	CTTGACTGGGAAGACGATTAC	50°C/30 s	~300bp	RT-PCR control
1042	UBI.R	*Ubiquitin*	CAGAAACCACCACGGAGACG			

Table 4 (continue). RT-PCR primers for genes involved in anthocyanin biosynthesis

2.4 Measurement of anthocyanin

2.4.1 Anthocyanin Extraction

For the measurement of anthocyanin in petal lobe tissue, the corolla of two flowers from each plant were weighed and incubated for 24 hours at room temperature in the dark in an extraction buffer consisting of 97:3 (v:v) methanol: concentrated (37%) HCl. Spectrography measurement was carried out as soon as possible afterwards to reduce the potential of anthocyanin degradation.

2.4.2 Spectrophotometer analysis

Absorption of the crude acidic methanol extract was measured between wavelengths of 190 nm -600 nm in 5 nanometer steps, using a 1 cM quartz glass cuvette in a scanning spectrophotometer.

2.5 Bioinformatics

2.5.1 Sequence analysis

Sequence data were obtained as AB1 trace files. Sequences were first aligned with Seqman 5.03 (DNASTAR Inc.) and alignments manually adjusted. Potential polymorphisms in aligned sequences were confirmed by inspection of the trace files. The alignments presented in this thesis were made with ClustalW (EMBL-EBI; www.ebi.ac.uk/Tools/msa/clustalw2/) developed by Larkin *et al*. (2007) and shaded manually.

2.5.2 BLAST searches

BLAST searches were carried out using the 'blastn' and 'blastx' commands from the NCBI website (http://www.ncbi.nlm.nih.gov/BLAST). Searches were limited to sequences of plant origin (Viridiplantae).

2.5.3 Linkage & QTL analysis

i. Linkage mapping

Map distance is used as a measure of genetic linkage, and reflects the frequency of crossovers between two loci. Because the presence of a crossover in one region reduces the likelihood of a crossover in an adjacent regions (crossover interference), and an even number of multiple crossovers between two loci are not detected, mapping functions are used to correct for these factors when deriving map distances from apparent recombination frequencies, for example by use of the Kosambi mapping function (Kosambi, 1944). This function corrects for moderate interference, according to the equation;

$$X = \tfrac{1}{4} \ln [(1+2\theta)/(1-2\theta)]$$

where X is the map distance and θ the recombination fraction. However, for loci that are closely linked, the correct map distance does not differ greatly from the recombination fraction.

Linkage analysis and map construction were done with JoinMap version 3.0 (Van Ooijen and Voorrips, 2001) with the Kosambi mapping function. A minimum LOD (logarithm of odds) score of 7 was used in assigning markers to a linkage group.

ii. QTL analysis

QTL analyses were carried out in GridQTL, a web-based programme developed at the University of Edinburgh (http://qtl.cap.ed.ac.uk/; Seaton *et al*., 2006). In the QTL mapping analysis, phenotypic values for the F_2 offspring were regressed onto indicator regressor variables for the additive (a) and dominance effects (d) of a suggestive QTL at fixed 1 cM intervals across the genome. The additive effect represents half of the average phenotypic difference between the two homozygous classes and dominance is the deviation of the mean of the heterozygous from the midpoint between the means of the two homozygote classes. Implementation of a chromosome-wide permutation test, of randomised phenotypes onto observed

genotypes, was used to set a threshold *F*-test score, above which QTL could be considered significant at $\alpha = 0.01$ or 0.05, preventing an excessive number of false positive QTLs being discovered. In this case for GridQTL, the permutations number was set as 1,000 (Vissher, Thomson and Haley, 1996). The most likely position of a QTL was taken to be the location giving the highest *F*-score. To overcome major QTLs preventing detection of smaller effect loci, tests were repeated after the largest-effect QTLs had been fixed as co-factors. This was repeated until at least seven significant QTLs were detected (there is a limit of six co-factors) or no more significant QTLs were detected at $\alpha = 0.05$. The estimated effect of substituting both alleles from one parent with the alleles from other parent (allele substitution effect) was normally calculated as followed. If $d \leq a$, the estimated effect will be 2a. In case of $d > a$, the effect will be $|a| + |d|$ with the sign (±) of a. The percentage of phenotypic variance (%Var) explained by each QTL was estimated was estimated as $[1-(MS_{full}/MS_{reduced})] \times 100$, where $MS_{full}/MS_{reduced}$ represents the residual variance with the QTL fitted.

2.5.4 Statistical analysis

Simple regression analysis and one-way *ANOVA* in this thesis were conducted using data-analysis add-in of Microsoft Excel® 2003.

Chapter 3

QTL analysis for flower colour and candidate genes

Flower colour variation had been investigated intensively since the early ages of genetics. Anthocyanin biosynthesis has been well-studied, both in term of biochemistry and genetics using mutations that have arisen during cultivation (Holton and Cornish, 1995; Tanaka *et al*. 2009; Martin *et al*., 1991, Springob *et al*. 2003). However, the basis for flower colour variation in nature has been more challenging. In some cases it has been possible to identify known structural genes or regulators of anthocyanin biosynthesis that are involved in natural variation, for example several loss-of-function mutations in *AN2*, encoding R2R3 MYB protein, were involved in the independent evolution of white flowers in *Petunia* species (Hoballah *et al*., 2007) or the *cis*-regulatory mutation that down-regulates F3'H causing flower hue difference, from purple-blue range to red, leading to the pollinator shift to hummingbirds in *Ipomea* family (Marais and Rausher, 2010). However, this has not been able to explain all variation found in nature, probably because many genes are involved in flower pigmentation and their interactions can be complex.

Quantitative trait locus (QTL) analysis is a first step to identify the genes or chromosome regions that are responsible for a complex trait, when the variation occurs between inter-fertile individuals. This chapter describes an experiment to identify QTL that are responsible for flower colour variation, involving anthocyanin pigmentation, between *Antirrhinum* species. This involved an existing F_2 hybrid population between *A. majus* and *A. charidemi*, called V1640. Having mapped QTL, candidate structural and regulatory genes were tested for their linkage to these QTL.

3.1 QTL analysis of the F_2 population

To identify chromosome region associated with flower colour variation in *Antirrhinum*. I used an F_2 population of *A. majus* x *A. charidemi* (V1640) consisting of 204 plants. *A. majus*, a member of subsection *Antirrhinum*, has darkly pigmented flowers, while the flowers of *A. charidemi* (subsection *Kickxiella*) are paler (Figure 4). F_2 plants had been grown together in a glasshouse and their flowers photographed by Dr Nicolas Langdale in 2002. I used flower images to score two traits independently; lobe colour (on an increasing scale of 1 to 10) and tube colour trait (on an increasing scale of 1 to 6; see Materials & Methods for details). The F_2 plants had also been genotyped at 169 loci and used to construct a linkage map consisting of eight linkage groups ranging from 70 to 104 cM, spanning 691 cM (Feng *et al.*, 2009).

The program GridQTL (Seaton *et al.*, 2006) was used to detect significant QTLs with the interval mapping method of Haley, Knott and Elsen (1994). The most likely position of a QTL was taken to be the location giving the highest LOD (log-of-odds) score and was considered significant when the peak exceeded the chromosome-wide significance threshold for $\alpha = 0.01$ or 0.05, as assessed by permutation of phenotypes onto genotypes in GridQTL. To identify minor-effect QTLs, the most likely QTLs were fixed as co-factors in further rounds of regression until no more significant QTL could be detected.

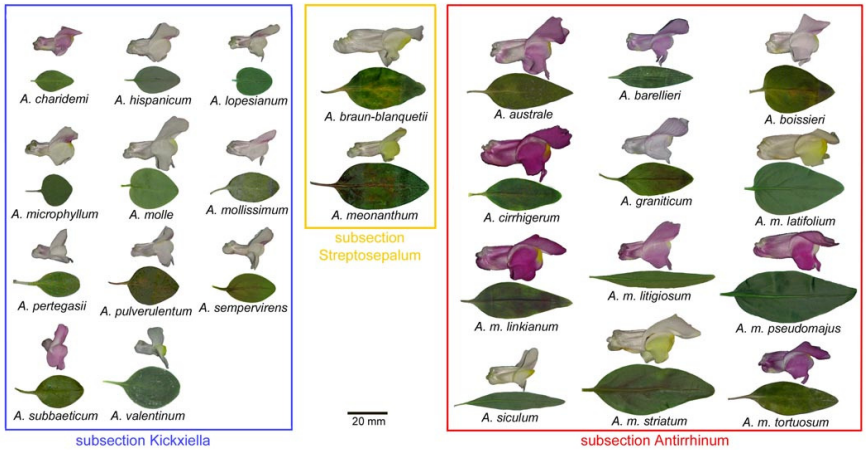


Figure 4. Flower colour variation within *Antirrhinum* species (Picture from Feng *et al.*, 2009).

3.1.1 QTL for petal lobe colour trait

For lobe colour, two significant QTL were detected at α= 0.01 when no loci were fixed as co-factors. The most likely QTL shows its highest LOD score at 59 cM in LG3 (Figure 7a) with a LOD score of 17.7, equivalent to an *F*-statistic of 50 (Table 8). The *A. majus* allele increased pigmentation (*A. majus* homozygotes scored on average 3.6 units higher than *A. charidemi* homozygotes) and heterozygotes had intermediate values (0.9 units higher than *A. charidemi* homozygotes). The second most-likely QTL was estimated at 46 cM in LG7 (Figure 5c), with a LOD score of 13.9 (*F* = 37.8). Again the *A. majus* allele increased pigmentation (*A. majus* homozygotes scoring 2.7 units higher than *A. charidemi* homozygotes) and showed incomplete dominance.

When the first QTLs (59 cM in LG3) was considered as genetic background and fixed as a co-factor, repeated analysis result showed two additional QTL with smaller effects. One mapped at 55 cM in LG5 (Figure 6a; Table 5) and the other at 28 cM in LG3.

Repeated-analysis with both major-effect QTL fixed as co-factors (60 cM LG3 and 45 cM LG7), revealed two more suggestive QTL at 7 cM in LG7 (Figure 5d) and 62 cM in LG2 (Figure 6b).

A final analysis with three most significant QTL fixed as co-factors, revealed another significant QTL, at 58 cM in LG8 (Figure 6c).

Together the QTL that exceeded the chromosome-wide *F*-statistic threshold when α= 0.01 could explain about 89% phenotype variation in lobe colour. The two most significant QTL (46 cM in LG7 and 59 cM in LG3) together could explain 60% of the variance (Table 5). All loci had alleles that acted in the parental direction (i.e., the allele from *A. majus* increased pigmentation).

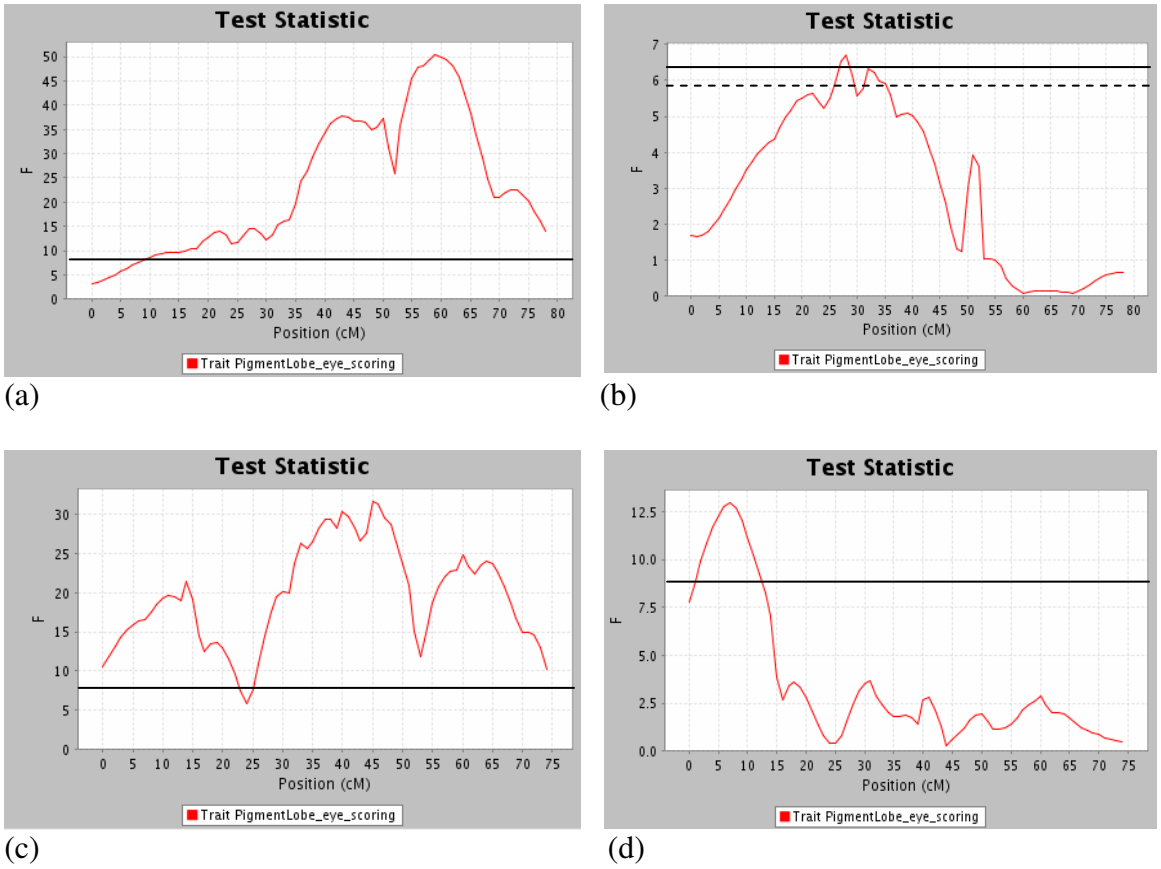


Figure 5. **Lobe colour QTL in LG 3 and LG 7** *F*-statistics are shown for LG 3 (a,b) and LG 7(c,d). The levels of chromosome-wide significance at α=0.05 is illustrated by a horizontal dashed line, at α=0.01 by a horizontal filled line. (a) In the first analysis, with no QTL fixed as co-factors, a QTL peak was detected at LG3 position 59 cM and an extra peak at 43 cM. (b) When the more likely QTL was fixed as genetic background, the extra QTL peak at position 43 cM was shifted to 26 cM with significance at α<0.01. (c) Including the effect of the LG3 60 cM QTL as genetic background, a QTL was detected in LG7 at position 46 cM and two extra peaks were present at 14 cM and 60 cM. (b) When QTLs at LG3 position 59 cM and LG7 position 45 cM were fixed, the peak at 60 cM disappeared and the peak at 14 cM was shifted to 7 cM with significance at α<0.01.

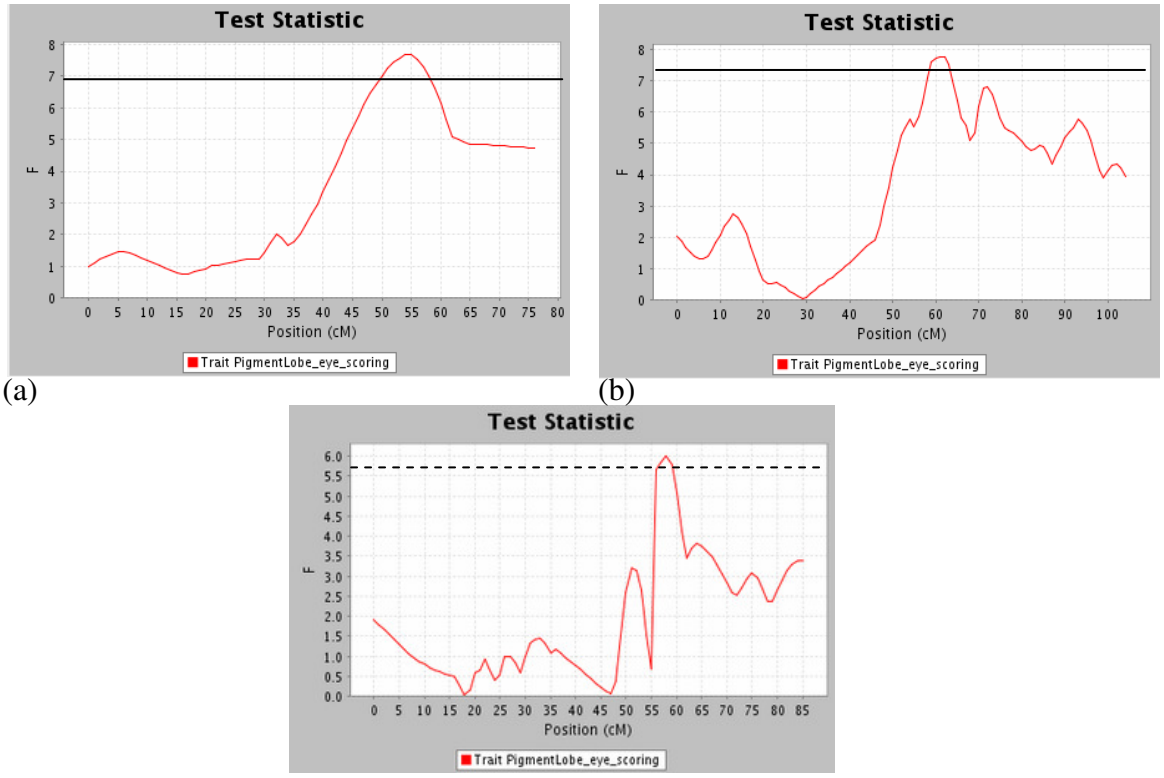


Figure 6. QTL for lobe colour in LG 5 and LG 7 (a) A peak in the F-statistic was detected in LG5 at 55 cM when LG3 position 60 cM was fixed as the genetic background. (b) a peak was detected at LG2 position 62 cM when both LG3 position 59 cM and LG7 position 45 cM were fixed. (c) A peak at LG8 at 56 cM when four QTL were fixed as background (LG2 position 54 cM, LG3 position 60 cM, LG7 positions 7 cM and 45 cM).

QTL position		Estimate				
LG	cM	*F*-statistic	LOD	a	d	%Var
3	59	50.45	17.71	1.78	0.861	33.09
7	45	37.78	13.90	1.33	-0.235	26.89
3	28	6.71	2.81	0.87	-0.016	5.45
5	55	7.69	3.22	0.75	-0.37	6.33
2	62	7.76	3.24	0.74	0.08	6.45
7	7	12.97	5.29	0.79	0.81	10.89

Table 5. Lobe colour QTLs Only QTL that are significant at $\alpha = 0.01$are shown. LG = Linkage group, cM = centimorgan, a = additive effect, d = dominance effect. These six QTLs together can explain 89% of phenotypic variation in the F_2 population.

3.1.2 QTL result for tube colour trait.

QTL analysis was performed for tube colour, as described for lobes. The first analysis yielded two major QTLs. The most significant, at 60 cM LG3 had an exceptionally high LOD score of 37 (Fig. 7a; Table 6). This position corresponds to one of the major QTL for lobe colour, suggesting that the gene affects pigmentation throughout the flower.

Extra peaks were also found in this region, as with lobe colour. However, these disappeared when the analysis was repeated with the major QTLs fixed as a cofactor (Figure 7b). The next most significant QTL was found in LG7 at position 46 cM (Figure 7c). Additional peaks were later found in LG7 when other QTL were fixed as background (Figure 7c, d). Two more significant QTL were detected in this round of analysis; in LG2 at position 88 cM (Figure 8b) and a smaller QTL, with marginal significance at LG 1 position 23 cM (Figure 8c).

When LG3 position 60 cM was considered to be genetic background, two more QTLs were detected. The more significant QTL was found in LG4 at position 12 cM (Figure 8a), the other at position 66 cM in LG7 (Figure 7d). Fixing these two QTL, in addition to the others, allowed detection of a minor peak at LG3 position 55 cM (Figure 10d). Repeating the analysis with LG3 position 60 cM, LG4 position 12 cM, LG2 position 89 cM and LG7 position 46 cM as the genetic background did not give any new suggestive QTL. However, with one extra cofactor (LG1 position 22cM), GridQTL revealed one more minor QTL, in LG7 position 8 cM (Figure 7d).

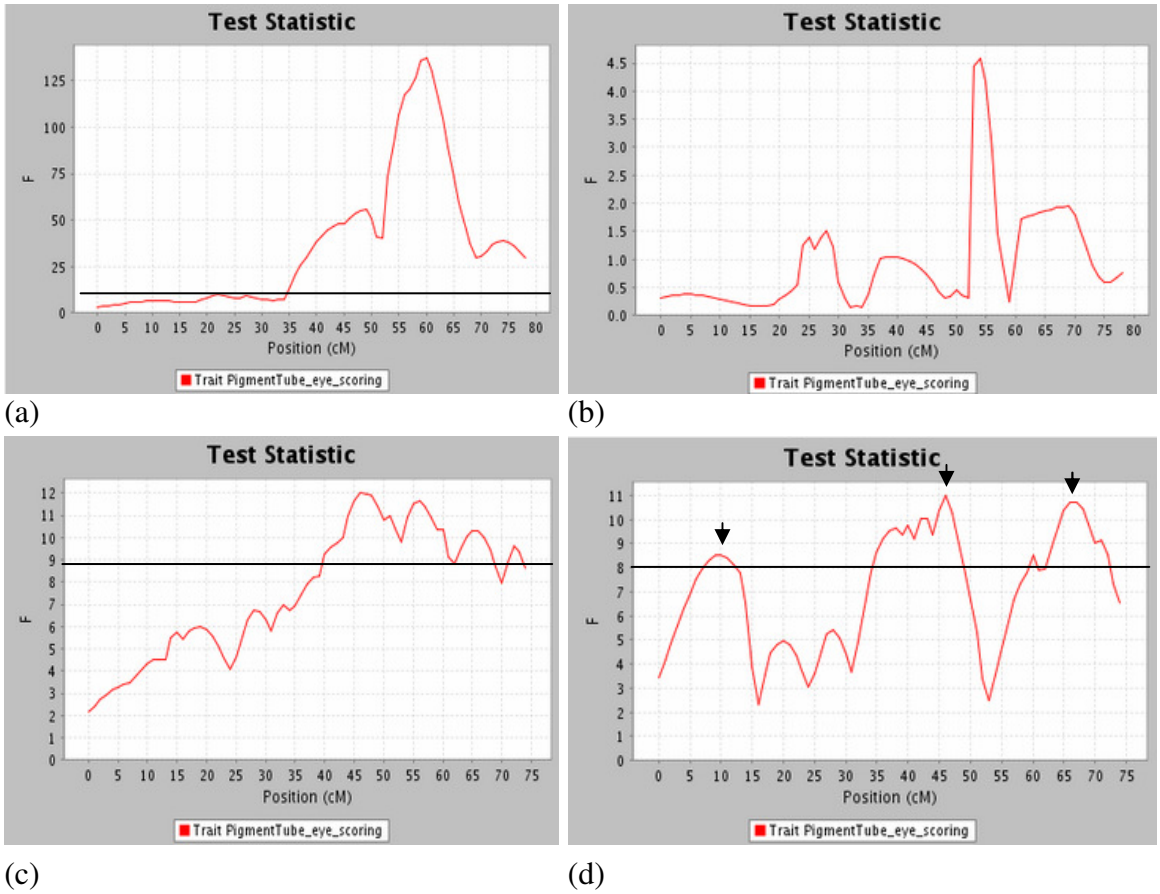


Figure 7. **Tube colour QTL in LG3 and LG7** *F*-statistics are shown for tube colour QTL in LG 3 (a,b) and LG 7(c,d). (a) The first analysis was done without genetic co-factors. The largest QTL peak was estimated at position 60 cM and an extra peak at 49 cM was detected. (b) Re-analysis with the QTL at 60 cM fixed as genetic background; the extra QTL peak has disappeared. (c) Another major QTL in LG7 position 46 cM was detected, with extra peaks at 56 cM and 66 cM. (d) When LG3 position 59 cM, LG4 position 12 cM and LG2 position 89 cM were fixed as genetic background, extra two minor QTLs were detect in LG 7 at 11 cM and 66cM with significance at α=0.01.

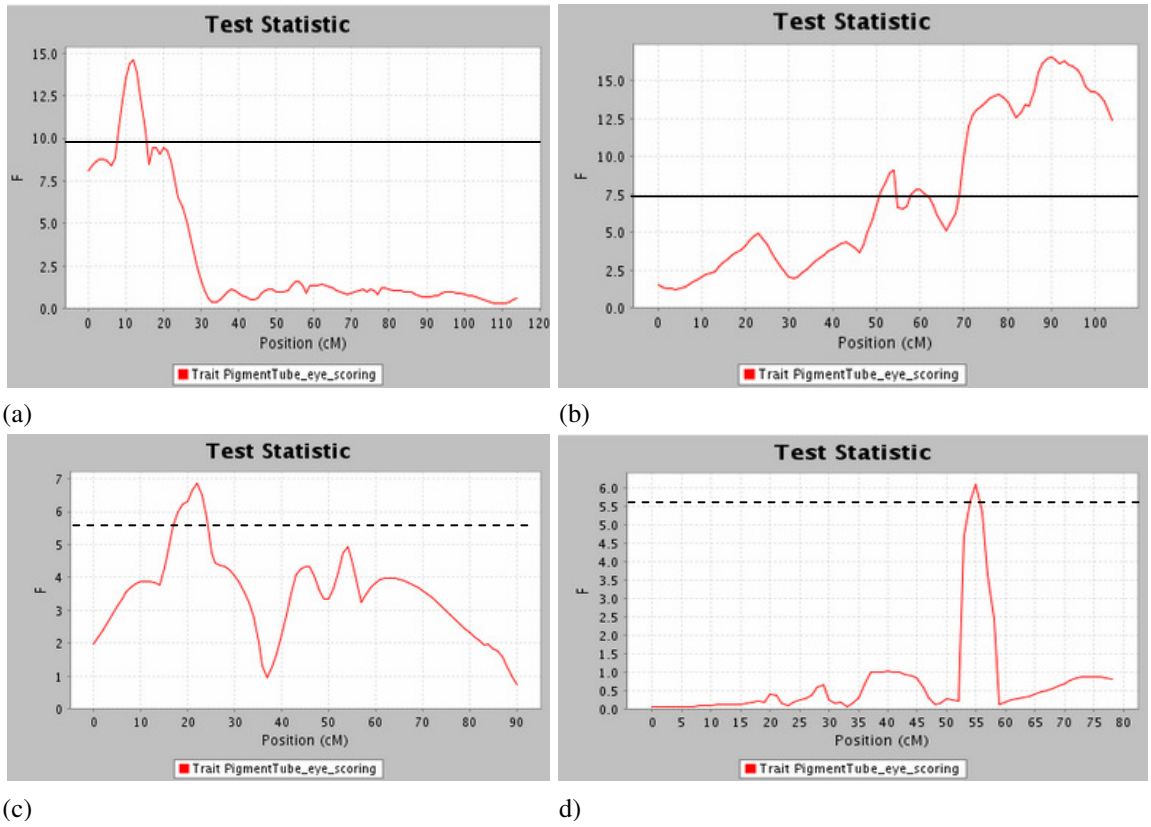


Figure 8. Tube colour traits in other chromosomes (a) A QTL was detected in LG4 at 12 cM, when LG3 position 60 cM was fixed as the genetic background. (b) A significant peak was detected at LG2 position 88 cM when LG3 position 59 cM was fixed as genetic background. (c) A significant peak was detected in LG1 at 22 cM when LG3 position 60 cM, LG7 position 46 cM and LG 2 position 88 cM were fixed. (d) A additional QTL in LG3 at 55 cM, was detected when LG3 position 60 cM, LG7 position 46 cM and LG 2 position 88 cM were fixed.

QTL position		Estimate				
LG	cM	F-value	LOD	a	d	%Var
3	60	137.3	37.41	1.493	0.602	57.68
7	46	12.03	4.93	0.654	0.027	9.93
2	88	8.76	3.65	0.534	-0.015	7.20
4	12	14.62	5.92	0.432	0.291	12.09
7	66	9.27	3.85	0.364	-0.040	7.79

Table 6. Tube colour QTLs Tube colour trait QTLs location (with significance at α = 0.01 and LOD score >3) and their details. LG = Linkage group, cM = centimorgan, a = additive effect, d = dominance effect. These five QTLs together can explain 90.23|% phenotypic variation in F_2.

In summary, two major QTLs and three lesser QTLs had *F*-statistics above the chromosome-wide expectation for α=0.01. Together these could explain more than 90% of the variance in tube colour in the *A. majus* x *A. charidemi* F_2 population (Table 6). All QTLs acted in the parental direction and showed incomplete dominance. The two major QTLs for tube colour (60 cM in LG3 and 46 cm in LG7) corresponded to the two major QTL for lobe colour, suggesting that these genes affect colour throughout the flower. However, the variance explained by the QTLs differed between the two tissues. In lobes the LG3 and LG7 QTL explained a similar proportion of the variance (33% and 27%, respectively), while the LG3 locus explained most of the variation in lobe colour (57%), compared to only 10% for the LG7 QTL. This suggests that the QTL might have different effects in different tissues. In addition, several QTLs were found that had significant effects only in tubes or lobes.

3.2 A molecular study of an LG3 candidate locus, *Rosea*

QTL analysis had detected a major-effect locus for both petal tube and lobe colour at around 60 cM in LG3. Additional *F*-statistic peaks were detected nearby when this QTL was fixed as a co-factor, suggesting involvement of more than one linked gene (a QTL-rich region). Fine mapping of complex traits have often yielded multiple constituent loci within a major QTL, e.g. Legare and Frankel, 2000; Demarest *et al*., 2001; Mozhui *et al*., 2008) and might reflect clustering of functionally related loci and genes (e.g., Petkov *et al*., 2006).

The QTL mapped close to the *Rosea (Ros)* gene, which had been reported to underlie differences in anthocyanin pigmentation between *Antirrhinum* species (Schwinn *et al*., 2006; Whibley *et al*., 2006). *Ros* is a regulatory gene which affects pattern and intensity of anthocyanin pigmentation in both lobe and tubes. Recently, *Ros* gene had been characterized molecularly which reveals that it is a complex locus consisting of two closely-linked R2R3 MYB genes, named *Ros1* and *Ros2*. The two genes are believed to have evolved from a recent gene duplication because of the very high similarity in their

coding regions. Analysis of mutants suggested that *Ros1* expression is necessary for high levels of pigmentation in flowers and that it might suppress *Ros2* expression. Though loss of *Ros2* activity decreased pigmentation, the effect was much weaker than for loss of *Ros1* (Schwinn *et al.*, 2006). *Ros* is therefore a candidate for the QTL at 60 cM in LG3.

The *Pallida* gene, which encodes dihydroflavonol reductase (DFR) enzyme, one of the key enzymes in anthocyanin biosynthesis pathway, was also found to locate about 16 cM further from *Rosea* and is therefore a second candidate gene (Luo *et al*., 1999).

3.2.1 Differences in *Rosea1* between *Antirrhinum* species

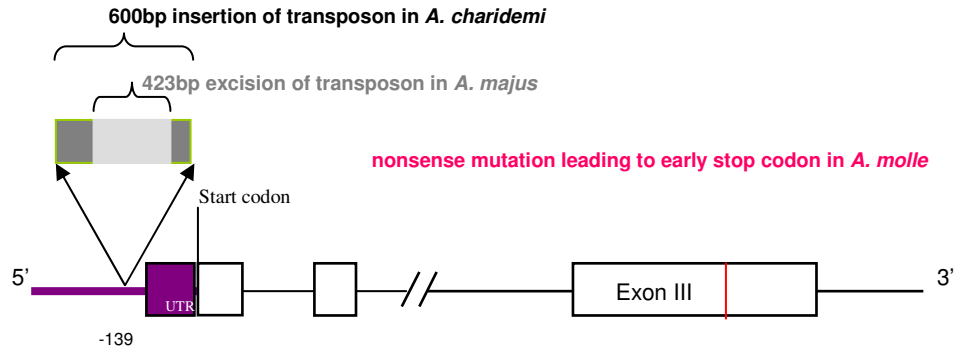
Sequences were obtained from the coding region, 5' upstream region and parts of the introns of *Ros1* from three *Antirrhinum* species (Figure 9a, detail in Table 2 for primer sequence and PCR condition). The coding regions were very well conserved. However, the *A. molle Ros1* allele (*AmoRos1*), which has white flowers, showed seven SNPs in the last exon compared to the other species with pink or dark flowers. One of these SNPs caused a frame-shift mutation, predicted to introduce an early stop codon at amino acid position 187 (Figure 10b). This is likely to cause a loss of function mutation in *Ros1*, suggesting that it might contribute to the white flowers of *A. molle*. This was tested in an F_2 hybrid family between *A. majus* and *A .molle*. All F_2 hybrids (n = 92) with white flowers (16 plants) were homozygous for the *A. molle Ros1* allele. Pigmented F_2 individuals from the same F_2 family were self pollinated to produce F_3 families. Though the colour in the progeny (n=36) still varied from pale to very dark, there was no white flower. This is consistent with the *AmoRos1* allele contributing to the white flowers of *A. molle* and the presence of other QTL involved in flower colour variation.

In the 5'upstream region of *Ros1*, an insertion-deletion was detected at position -139 relative to the initiation codon. The *A. majus Ros1* allele (*AmRos1*) had an additional 177 bp and the *A. charidemi* allele *(AcRos1)* at least 600 bp relatively to *AmoRos1*. A

database BLAST search showed that the big 0.6kb insertion was 75% similar to *Tam661*, one of the transposon elements in *Antirrhinum* (accession number AB012941, complete sequence length 3692 bp), suggesting that it originated from a transposon insertion. The smaller 177 bp insertion in the *AmRos1* allele corresponded to loss of 423 bp from the middle of the transposon, suggesting that it had arisen by internal deletion of the transposon present in the *AcRos1* allele. Insertions of Type II transposons like Tam661 are usually found to be flanked by a small direct duplication of target sequence generated at the time of insertion. Also, the excision of this class of transposon is almost always associated with a sequence footprint (reviewed by Colot, Haedens and Rossignol, 1998). However, the Tam661 trace found in 5'UTR of *Ros1* showed no clear evidence of insertion or deletion causing the differences between *AmRos1* and *AcRos1* alleles. The *AmoRos1* alleles showed no evidence of any footprint, suggesting that it was derived from the ancestral allele before transposon insertion.

The differences found in the 5' upstream and promoter region in *Ros1* from different species made it plausible that these polymorphisms are major determinants of *Ros1* expression and might contribute to flower colour variation. Polymorphisms in the promoter region were also suggested by Schwinn *et al*. (2006) to explain why the $Ros^{dorsalis}$ (Ros^{dor}) allele, which carries a deletion around -187, showed reduced activity in some parts of the flower. In an attempt to find the regulation region, Pathirana (2009) used promoter deletion analysis to identify a region in *Ros1* between position -162 and -120 which was predicted to be required for activating expression. However, her result showed no significant association between allelic expressions of chimeric *Ros1* in *A. majus* variety and flower pigmentation. According to Pathirana, the promoter region of *Ros1* should be in the region between position -192 and -109 of the *A. majus* ^{dor} *Ros1* allele. Comparing to the sequences analysed in this project [*AmRos1* (JI7, JI.98) *AcRos1* and *AmoRos1*], the promoter region should be located between -178 and – 95 (Figure 9b). Comparing the 5'UTR and promoter regions in this experiment, an extra WRKY recognition site (TTGACY; Ciolkowski *et al*.,2008) were created by the incomplete *Tam661* transposon in *AmRos1* at position -176, which can possibly enhance the

(a)



(b)

```
AmoRos1              CTACCTAC-TGTATCTTGGAAATATCTGCTTCATGGGAGCCAAAaaaGGGGgTGACAAG -135
AmRos1              TCTACCTACCTGTATCTTGGAAATATCTGCTTTATGAGAGCCAAA---GGGG-TGACAGG -331
AcRos1              ------------------------------------------------------------ -777

AmoRos1              ------------------------------------------------------------ -135
AmRos1              TTATAGTA-GGGTGTTCGCTGGCGGATTTGGGCGGGTTGTCGTTAAAAATATGACCATAC -271
AcRos1              ---------GGGTGTTCGCTGGCGGATTTGGGCGGGTTGTCGTTAAAAATATGACCATAC -726
Tam 661             --ATAATAGGGGTGTTCGCGGGCGGGTTTAGGCGGGTTGTCGTTAAAAATGTGACCATAC 58

AmoRos1             ------------------------------------------------------------ -135
AmRos1              TCAAATTTTGCGGGTTACCAATTTTTCAACCCG---------AGTTAAAAGCTAAACCAT -211
AcRos1              TCAAATTTTGCGGGTTACCAATTTTTCAACCCG---------AGTTAAAAGCTAAACCAT -675
Tam 661             CCAAACTTTGCGGGTTACCACTTTTTCAACCCGGGCGGGTTAAGTTAAAAACTAAACCAT 118

AmoRos1             ------------------------------------------------------------ -135
AmRos1              GCGGGTTGCGGGTTGGGCG------------------------------------------ -192
AcRos1              GCGGGTTGCGGGTTGGGCGGGTTGAGCGGGTGGCGGGTTTAAATTATGGAAAAAATATAG -615
Tam 661             GCGGGTTGCGGGTTGGGCGGGTTGAGCAGGTGGCGGGTTTAAATTATGG-GAAAATATAG 177

AmoRos1             ------------------------------------------------------------ -135
AmRos1              ------------------------------------------------------------ -192
AcRos1              TAATATTTTGAAATAAAAACAAAAACATAATTTACTTAGCATTCTAAATAACAACAAAAA -555
Tam 661             TAATATTTCGGTATAAAAACAAAAACATAATTTACTTTGCGTTCTAAAAAACAACAAAAA 178

AmoRos1             ------------------------------------------------------------ -135
AmRos1              ------------------------------------------------------------ -192
AcRos1              TAATACCATAACTTAATAAAAATACTTAACCACATCATAACTTAATTAAAATAACAATAT -495
Tam 661             TAATACAATAACTTAATAAAAACACTTAACCACAACATAACTTAATTAAAATAACAATAT 238

AmoRos1             ------------------------------------------------------------ -135
AmRos1              ------------------------------------------------------------ -192
AcRos1              AGGAATATAAACGAATATTTTATCCTCATAAATTATAAAATAACAATTTAAATCGATTCA -435
Tam 661             AAGAATGTAAACGAATATTTTTTTGCATAAATTATAAA-TTACAATTTAAATCGATTTA 297
```

Figure 9. Structure of the *Rosea1* gene and its variation between species. (a) Schematic allelic structure is shown for the alleles from *A. majus*, *A. charidemi* and *A. molle* (not to scale). The boxes indicate exon and UTR, lines indicate non-coding regions and red vertical lines indicate SNPs. The dark and light gray boxes together indicated the 600 bp *Tam661* in *A. charidemi* with its dark grey portion showing the 177 bp residue of *Tam661* left in *A. majus*. The red line in the third exon indicated the SNPs that terminates the open reading frame in *A. molle*. The thick purple line is the 5'UTR as detailed in (b) showing the Tam661 sequence trace in 5'UTR of *Ros1* allele between *Antirrhinum* species. (To be continued.)

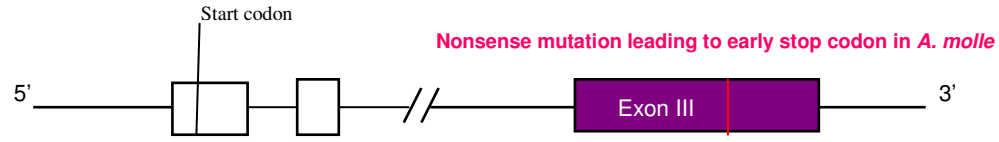
(b; continued)

```
AmoRos1          ------------------------------------------------------------ -135
AmRos1           ------------------------------------------------------------ -192
AcRos1           TAATTCAAAATACAACTATCAAATAACATAAAATTAACAAAAAAAACTATATTATAGCAA -375
Tam 661          TATTTCAATA---------------------------------------------------- 307

AmoRos1          ------------------------------------------------------------ -135
AmRos1           ------------------------------------------------------------ -192
AcRos1           TATTATGAACATACAATTATATGATTTCTGAAAACTTTCCTTGCTCCTCTCATTCCTCTA -315
Tam 661          --TTATGAACATACAATTATATGACTTCCGGGAACTTTCCTTGCTCCACTCAATCCTCCA 315

AmoRos1          ------------------------------------------------------------ -135
AmRos1           ------------------------------------------------------------ -192
AcRos1           -AAATATACAAACTAAGGATGTATGCGGGCGGGTTCGGGCGGATTGTCGTTAAAAATTAG -255
Tam 661          CAAATATTAAAACTAGGGATGTATGCCGGCGGGTTCGGGCGGATTGTCCTTTAAAATGTG 375

AmoRos1          ------------------------------------------------------------ -135
AmRos1           ------------------------------------------------------------ -192
AcRos1           ACCATATCCAAGATTTGTGGGTTACCAATTTGTCAACCCGAGCGGGTCAAGTTAAAAACT -195
Tam 661          ACCATACGCAAAATTTATGGGTTACCAATTTGTCAATCCGAGGGGGTTAAGTTAAAAACT 435

                                                 WRKY recognition site
AmoRos1          ------------------------------------------------------------ -135
AmRos1           -------------------------AGTCGGGTTTTGACGGGTCTGGCGGATCG-AACG -159
AcRos1           AAACCAT-CGGGTT-CGG---CGG--A-TCGGG---------------CGGATCGGA-CG -159
RosdorP          ----------------------GGGCGAGTCGGGTTTTGACGGGTCTGGCGGATCGAAACG -153
Tam 661          AAACTATGCGGGTTGCGGGTTCGGCAAGTCGGGTTTTGA---------CGGATCG-GGCG 485

AmoRos1          -------------------CTACAGCCATAAAAAGGCCTATTTAAACCCGTGAAAGTTTC -94
AmRos1           GGTCATGAACACCCCTAGGCTACAGCCATAAAAAGGCCTATTTAAACCCGTGAAAGTTTC -99
AcRos1           GGTCATGAACACCCCTAGGCTACAGCCATAAAAAGGCCTATTTAAACCCGTGAAAGTTTC -99
RosdorP          GGTCATGAACACCCCTAGGCTACAGCCATAAAAAGCC-TATTTAAACCCGTGAAAGTTTC -94
Tam 661          GGTCATGAACACCCCTA 502
                                                                        W box
AmoRos1          GCTCAAGGGGTACTCATTAAAAAAAAGGGAAAGAGCAGCTAGACATGTGTTTTCTGTTTT -34
AmRos1           GCTCAAGGGGTACTCATTAAAAAAA-GGGAAAGAGCAGCTAGACATGTGTTTTCTGTTTT -39
AcRos1           GCTCAAGGGGTACTCATTAAAAAAA-GGGAAAGAGCAGCTAGACATGTGTTTTCTGTTTT -39
RosdorP          GCTCAAGGGGTACTCATTAAAAAAA-GGGAAAGAGCAGCTAGACATGTGTTTTCTGTTTT -54

                                                    Start codon
AmoRos1          GACACTTTTAACGAACGGGCATAGTA--TA--AAACGCAatg 3
AmRos1           GACACTTTTAACGAACGGGCATAGTACGTATTAAACGCAatg 3
AcRos1           GACACTTTTAACGAACGGGCATAGTACGTATTAAACGCAatg 3
RosdorP          GACACTTTTAACGAACGGGCATAGTA----TTAAACGCAatg 3
```

Figure 9 (continued). Structure of the *Rosea1* gene and its variation between species. (b; continued) the Tam661 sequence trace in 5'UTR of *Ros1* allele between *Antirrhinum* species. (RosdorP: Ros1 promoter sequence from Pathrirana, 2009)

expression of *AmRos1* (Figure 9b), as it is in the promoter region. WRKY transcription factors are known to regulate the regulatory genes in *Arabidopsis* (reviewed by van Verk, Bol and Linthorst, 2011). However, the only report that shows an association between anthocyanin regulatory gene and a WRKY transcription factor is the work by Tohge *et al*. (2005). Their experiments showed that the *TTG2* gene encodes a WRKY transcription factor, that is up-regulated when *PAP1* is over-expressed (Tohge *et al*.

(a)




```
AcRos1      TTGAACACAAAAACCATTAAGCAGACTAATATCGTAAGACCCCGAGCTCGGACCTTCACC 60
AmoRos1     TTGAACACAAAAACCATTAAGCAGACTAATATCGTAAGACCCCGAGCTCGGACCTTCACC 60
AmRos1      TTGAACACAAAAACCATTAAGCAGACTAATATCGTAAGACCCCGAGCTCGGACCTTCACC 60

AcRos1      GGATTGCACGTTACTTGGCCGAGAGAAGTCGGAAAAACCGATGAATTTTCAAATGTCCGG 120
AmoRos1     GGATTGCACGTTACTTGGCCGAGAGAAGTCGGAAAAACCGATGAATTTTCAAATGTTCGG 120
AmRos1      GGATTGCACGTTACTTGGCCGAGAGAAGTCGGAAAAACCGATGAATTTTCAAATGTCCGG 120

AcRos1      TTAACAACTGATGAGATTCCAGATTGTGAGAAGCAAACGCAATTTTACAATGATGTTGCG 180
AmoRos1     TTAACAACTGATGAGATTCCAGATTGTGAGACGCAAAAGCAATTTTACAATGATGTTGTG 180
AmRos1      TTAACAACTGATGAGATTCCAGATTGTGAGAAGCAAACGCAATTTTACAATGATGTTGAG 180

                              Early stop codon
AcRos1      TCGCCACAAGATGAAGTTGAAGACTGCATTCAGTGGTGGAGTAAGTTGCTAGAAACAACG 240
AmoRos1     TCGCCACAAGATGAAGTTGATGAATGCATTCAGTGGTGGAGTAAGTTGCTAGAAACAACG 240
AmRos1      TCGCCACAAGATGAAGTTGAAGACTGCATTCAGTGGTGGAGTAAGTTGCTAGAAACAACG 240

AcRos1      GAGGATGGGGAATTAGGAAACCTATTCGAGGAGGCCCAACAAATTGGAAATTAA 293
AmoRos1     GAGGATGGAGAATTAGGAAACCTATTCGAGGAGGCCCAACAAATTGGAAATTAA 293
AmRos1      GAGGATGGGGAATTAGGAAACCTATTCGAGGAGGCCCAACAAATTGGAAATTAA 293
```

(b)

```
AmRos1      MEKNCRGVRKGTWTKEEDTLLRQCIEEYGEGKWHQVPHRAGLNRCRKSCRLRWLNYLRPN 60
AmoRos1     MEKNCRGVRKGTWTKEEDTLLRQCIEEYGEGKWHQVPHRAGLNRCRKSCRLRWLNYLRPN 60

AmRos1      IKRGRFSRDEVDLIVRLHKLLGNKWSLIAGRIPGRTANDVKNFWNTHVGKNLGEDGERCR 120
AmoRos1     IKRGRFSRDEVDLIVRLHKLLGNKWSLIAGRIPGRTANDVKNFWNTHVGKNLGEDGERCR 120

AmRos1      KNVMNTKTIKLTNIVRPRARTFTGLHVTWPREVGKTDEFSNVRLTTDEIPDCEKQTQFYN 180
AmoRos1     KNVMNTKTIKQTNIVRPRARTFTGLHVTWPREVGKTDEFSNVRLTTDEIPDCETQNNFTM 180

AmRos1      DVASPQDEVEDCIQWWSKLLETTEDGELGNLFEEAQQIGN* 220
AmoRos1     MLFATR* 187
```

Figure 10. ***Rosea*** **allelic difference in last exon leading to premature stop codon in** ***A. molle*** (a) Third exon (purple box) sequence comparison of *A. majus*, *A. charidemi* and *A. molle*. There is no SNP between *AmRos1* and *AcRos1* in the third exon, though these differ for SNPs from the *AmoRos1*. The nonsense mutation in *A.moRos1* allele leads to early termination of amino acid sequence as shown in (b) Amino acid alignment of predicted *Rosea* protein in *AmRos1* and *AmoRos1*. The amino acid sequences diverge from amino acid position 174, with an early stop codon at amino acid position 187.

2005). However, this does not directly support the suggestion that the WRKY recognition site in the *AmRos1* promoter is involved in differences in *Ros1* regulation.

The species were also likely to differ in the *Ros2* gene, which is partially deleted in *A. majus* but apparently intact and expressed in *A. charidemi* (Schwinn *et al*., 2006; Manuela Costa, unpublished data). However, the genes are very closely linked, so *Ros1* can be use as the marker to test involvement of both genes in QTL analysis.

3.3 A molecular study of an LG3 candidate locus, *Pallida*

As being reported to be linked to *Ros* (Luo *et al*., 1999). I cannot test the effect of *Pallida* to flower pigmentation in V1640 population without the effect of *Rosea*. The test of *Pallida* was carried out in a near isogenic line (NIL) population which would be discussed in the next chapter, as well as the molecular study.

3.4 Sequence variation in the LG7 candidate, *Incolorata*

The QTL with second highest LOD score for both lobe and tube colour was mapped to 46 cM in LG7. Chromosome 7 also had been previous reported as the location of the *Incolorata* (*Inc*) gene which encodes flavanone-3-hydroxylase (F3H), one of the key enzymes of the anthocyanin biosynthetic pathway, catalyzing hydroxylation of flavanones (naringenin or eriodictyol) on the 3 position to form dihydroflavonols (Martin and Gerats, 1993). This enzyme was also believed to catalyse the rate-limiting step of anthocyanin biosynthesis (Chaffe, 2002). Consistent with this, wild-type alleles are not fully dominant over loss-of-function mutations (Stubbe, 1960). *Inc* was therefore a candidate for the QTL in LG7 and its structure was investigated further.

3.4.1 *Incolorata* allelic comparison between *Antirrhinum* sp.

Inc sequences were obtained from the 5' region, first exon, second intron and about 425 bp of the last exon of *Inc* from *A. majus* (*AmInc*), *A. molle* (*AmoInc*) and *A. charidemi* (*AcInc*). The 5' region was quite strongly conserved between alleles (Figure 12). However, a microsatellite around position – 31 from the ATG had been reported in the published *Inc* sequence from *A. majus* line JI.522 (S60876), but was missing in *AmInc* (JI.98), *AmoInc* and *AcInc* in this study.

In the first exon only three SNPs were detected between alleles (Figure 15a), at position +50, +192 and +193; the second and the third SNPs together causes an amino acid substitution at position 64 from Glycine (G) in *AmInc* to Lysine (K) in *AmoInc* and Glutamic acid (E) in *AcInc* (Figure 13b).

On the other hand, several SNPs and several stretches of sequence insertions and deletions (indels) were found in the second intron, so that this intron differed in size between species (*AmInc* 1379 bp, *AmoInc* 1327 bp and *AcInc* 1417 bp, data not shown). Interestingly, this second intron also found to contain duplicated imperfect tandem repeats of 527 to 562 bp which showed 81% similarity (Figure 14) reflecting their origin through duplication. This repeat can possibly have association with the expression of *Inc*, as in some cases found in human. One of the examples is in human disease, *Friedreich ataxia* disorder which patients carried the *FRDA* allele, encoding a protein called frataxin, which had numbers of short (GAA) tandem repeats in the first intron much higher than normal allele (Filla *et al*., 1996; reviewed by Gemayel *et al*., 2010). More of the association with tandem repeats in introns with expression also found in longer repeats, such as the work of Bellizzi *et al*. in 2005 showing the increasing copy numbers of a 72 bp repeat in the fifth intron of *SIRT3* (the mammalian ortholog of yeast silencing and longevity gene *SIR2*) can change the gene expression (reviewed by Gemayel *et al*., 2010).

The last exon and 3'downstream region were found to be more highly conserved, with only six SNPs differentiating the *AmInc* and *AcInc* alleles. However, the *AmoInc* allele revealed more variation with twelve SNPs comparing to *A. majus* in the same region. In the coding region, only one synonymous SNP found between *AmInc* and *AcInc* whereas seven SNPs were found between *AmInc/AcInc* and *AmoInc*; Six of them were synonymous but one, causes a replacement polymorphism which produces premature stop codon making the peptide 42 amino acids shorter (Figure 15b).

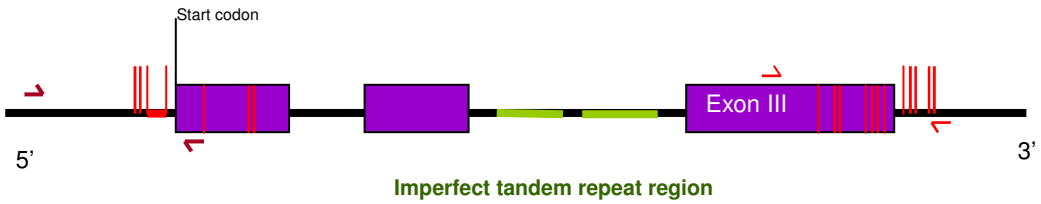
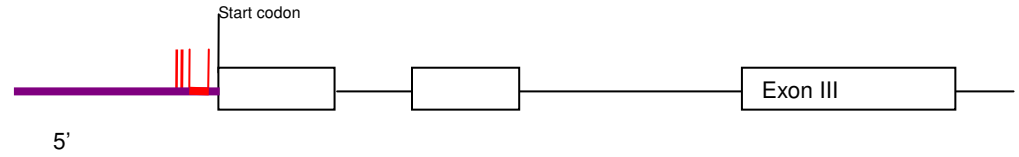


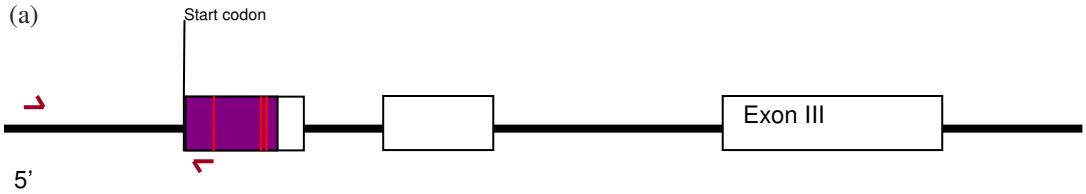
Figure 11. *Inc* alleles from different *Antirrhinum* species Red horizontal line in 5'UTR indicates the various-size missing microsatellite. The green line showed rough position of imperfect tandem repeats in second intron. Longer intron in *A. majus* (1379 bp) and *A. charidemi* (1417 bp) comparing to *A. molle* allele (1327 bp) were resulting from scattered indels in second intron of each allele.




```
AcInc          CCGTTAAGATATGAATTTAGAGTACTTAATACCGTGTATCCTCCTTATTCTTGTTATTAT 60
AmoInc         CCGTTAAGATATGAATTTAGAGTACTTAATACCGTGTATCCTCCTTATTCTTGTTATTAT 60
AmInc          CCGTTAAGATATGAATTTAGAGTACTTAATACCGTGTATCCTCCTTATTCTTGTTATTAT 60

AcInc          GGGAGTGGTCTAGCTTTAAGGTACGAATAAACCACAGGCATAAATTAACCACGATTCAAT 120
AmoInc         GGGAGTGGTCTAGCTTTAAGGTACGAATAAACCACAGGCATAAATTAACCACGATTCAAT 120
AmInc          GGGAGTGGTCTAGCTTTAAGGTACGAATAAACCACAGGCATAAATTAACCACGATTCAAT 120

AcInc          CAACTATATTATACCACTATATTTCTTCAAATGACATGTCAAAATTATATTTACAACCCT 180
AmoInc         CAACTATATTATACCACTATATTTCTTCAAATGACATGTCAAAATTATATTTACAACCCT 180
AmInc          CAACTATATTATACCACTATATTTCTTCAAATGACATGTCAAAATTATATTTACAACCCT 180

AcInc          AGTGACTTTTTCAAAATTACTACGTAAACCAATCTATATAATTTATATAACTATAACCTC 240
AmoInc         AGTGACTTTTTCAAAATTACTACGTAAACCAATCTATATAATTTATATAACTATAACCTC 240
AmInc          AGTGACTTTTTCAAAATTACTACGTAAACCAATCTATATAATTTATATAACTATAACCTC 240

AcInc          ATTTCGATTCTCTTGGTATGCCAGTTTCTTAGAAAATATACCAAAAGTATAACAATTTAT 300
AmoInc         ATTTCGATTCTCTTGGTATGCCAGTTTCTTAGAAAATATACCAAAAGTATAACAATTTAT 300
AmInc          ATTTCGATTCTCTTGGTATGCCAGTTTCTTAGAAAATATACCAAAAGTATAACAATTTAT 300
S60876         ----------------------------------------------AAGTATAACAATTTAT 16

AcInc          TTATATTGCAAGACATTAATTACGACACTA--AGACTAGCGAGTTATTTATAGACACGAA 358
AmoInc         TTATATTGCAAGACATTAATTACGACACTA--AGACTAGCGAGTTATTTATAGACACGAA 358
AmInc          TTATATTGCAAGACATTAATTACGACACTA--AGACTAGCGAGTTATTTATAGACACGAA 358
S60876         TTCTATTGCAAGACAACAATTACTAAACTATAAGACTAGCGAATTATTTATAGACACGAA 76

AcInc          GAAAAGGCTATCACAAATATACTT-TTAAAGTAATTAAATTTTATTTAAAGAACGGGCAA 417
AmoInc         GAAAAGGCTATCACAAATATACTT-TTAAAGTAATTAAATTTTATTTAAAGAACGGGCAA 417
AmInc          GAAAAGGCTATCACAAATATACTT-TTAAAGTAATTAAATTTTATTTAAAGAACGGGCAA 417
S60876         AAAAAGGCTGTCACGAATATACTTCTTAAAGCAATTAAATTTTATTAAAAGAACGG-CAA 135

AcInc          TGCATGACAGAATTTAAAATATCATTTTCCTAAAGATTATATATGAAGTTGAAAGTTTAG 477
AmoInc         TGCATGACAGAATTTAAAATATCATTTTCCTAAAGATTATATATGAAGTTGAAAGTTTAG 477
AmInc          TGCATGACAGAATTTAAAATATCATTTTCCTAAAGATTATATATGAAGTTGAAAGTTTAG 477
S60876         ---ATGACAGAATTTAAAATGTCATTTTCCTAAAGATTATATATGAAGTTGAAAGTTTAG 192

AcInc          TACCAGGAAGGTTAGGTGCGACGACGGTTGATACAAAGCTACTCTTCACACCTCACGCAC 537
AmoInc         TACCAGGAAGGTTAGGTGCGACGACGGTTGATACAAAGCTACTCTTCACACCTCACGCAC 537
AmInc          TACCAGGAAGGTTAGGTGCGACGACGGTTGATACAAAGCTACTCTTCACACCTCACGCAC 537
S60876         TAACAGGAAGGTTAGGTGCGACGACGGTTGATACAA-GCTACTCTTCACACCTCACGCAC 251

AcInc          CCCTCCGCAATCTCGCTATATAAATTTCCGCGAAACTTATGAAATTCAAAATAGCACCAG 597
AmoInc         CCCTCCGCAATCTCGCTATATAAATTTCCGCGAAACTTATGAAATTCAAAATAGCACCAG 597
AmInc          CCCTCCGCAATCTCGCTATATAAATTTCCGCGAAACTTATGAAATTCAAAATAGCACCAG 597
S60876         CCCTCCGCAATCTCGCTATATAAATTTCCGCAAAACTT-TGAAATTCAAA-TAGTACCAA 309

AcInc          CACAAATTTCCATAGTACAAAAAGAAAA-------GAAAAGAAAAGAAAAGAGAAATCAA 650
AmoInc         CACAAATTTCCATAA-AAAAAAAGAAA--------G-AAAGAAAAGAAAAGAGAAATCAA 647
AmInc          CACAAATTTCCATAGTACAAAAA-AA---------GAAAAGAAAGGAAAAGAGAAATCAA 647
S60876         CACAAATT-CCATAGTACAAAAAAAAAAAAAAAAAAAAGAAAAGAAAAGAGAAATCAA 368

AcInc          TCAAATG 657
AmoInc         TCAAATG 654
AmInc          TCAAATG 654
S60876         TCAAATG 375
```

Figure 12. 5' *Inc* polymorphisms between species S60876 is the *Inc* allele from *A. majus*, line JI:522 in NCBI (accession number S60876). Indels, SNPs and size variation at a microsatellite were found in the *A. majus* allele sequenced here comparing to the published sequence. In addition, three SNPs was found between *AmInc* and *AmoInc*, and 1 SNP between *AmInc* and *AcInc*.

(a)

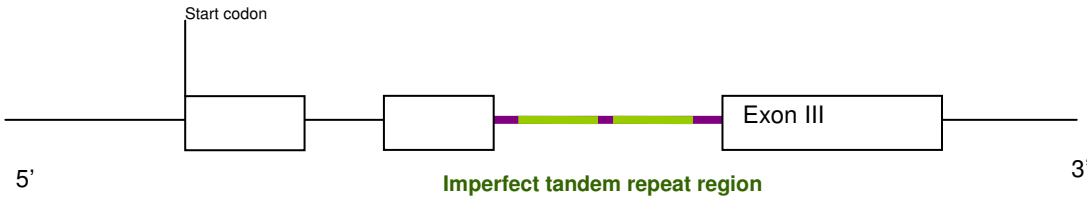



```
AcInc          ATGAATTCAGCCACGATATTCCGGTTATATCACTAGCTGGAATAGATGGTGATGATAGGG 60
AmoInc         ATGAATTCAGCCACGATATTCCGGTTATATCACTAGCTGGAATAGATGGTGATGATAGGG 60
AmInc          ATGAATTCAGCCACGATATTCCGGTTATATCACTAGCTGGAATAGATGGCGATGATAGGG 60

AcInc          TGGAGATTTGTCGGAAAATAGTGGCGGCGTGCGAAGAATGGGGTATTTTTCAGGTTGTTG 120
AmoInc         TGGAGATTTGTCGGAAAATAGTGGCGGCGTGCGAAGAATGGGGTATTTTTCAGGTTGTTG 120
AmInc          TGGAGATTTGTCGGAAAATAGTGGCGGCGTGCGAAGAATGGGGTATTTTTCAGGTTGTTG 120

AcInc          ATCATAGTATTGATTTGAAGATTGTTAGCGAAATGACCGATATGGCACGAGAGTTCTTCG 180
AmoInc         ATCATAGTATTGATTTGAAGATTGTTAGCGAAATGACCGATATGGCACGAGAGTTCTTCG 180
AmInc          ATCATAGTATTGATTTGAAGATTGTTAGCGAAATGACCGATATGGCACGAGAGTTCTTCG 180

AcInc          CTATGCCTGCGGAAGAGAAGCTTCGTTTCGACATGTCGGGAGGGAAGAAAGGTGGTTTCA 240
AmoInc         CTATGCCTGCGAAAGAGAAGCTTCGTTTCGACATGTCGGGAGGGAAGAAAGGTGGTTTCA 240
AmInc          CTATGCCTGCGGGAGAGAAGCTTCGTTTCGACATGTCGGGAGGGAAGAAAGGTGGTTTCA 240

AcInc          TCGTCTCTAGCCATCTTCAGA 261
AmoInc         TCGTCTCTAGCCATCTTCAGA 261
AmInc          TCGTCTCTAGCCATCTTCAGA 261
```

(b)

```
AmInc          EFSHDIPVISLAGIDGDDRVEICRKIVAACEEWGIFQVVDHSIDLKIVSEMTDMAREFFA 60
AmoInc         EFSHDIPVISLAGIDGDDRVEICRKIVAACEEWGIFQVVDHSIDLKIVSEMTDMAREFFA 60
AcInc          EFSHDIPVISLAGIDGDDRVEICRKIVAACEEWGIFQVVDHSIDLKIVSEMTDMAREFFA 60

AmInc          MPAGEKLRFDMSGGKKGGFIVSSHLQX 87
AmoInc         MPAKEKLRFDMSGGKKGGFIVSSHLQX 87
AcInc          MPAEEKLRFDMSGGKKGGFIVSSHLQX 87
```

Figure 13. Allelic variation in first exon of *Inc* (a) Partial sequence of the first exon, with 3 SNPs found at position +50, +192 and +193, (b) making the substitution of in the predicted amino acid sequence at position 64 from Glycine (G) in *AmInc* to Lysine (K) in *AmoInc* and Glutamic acid (E) in *AcInc*.

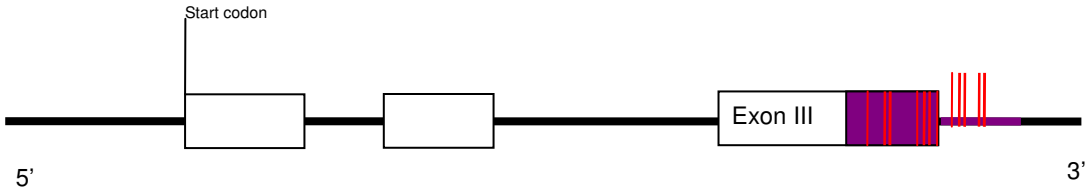


Score = 517 bits (572), Expect = 4e-150, Identities = 462/572 (81%), Gaps = 53/572 (9%)

```
First repeat   53    AAAATTACTGTTGGCTGTTAACACTACGTTATTCA----TGAAGAAACTTAAAGATTCCA  108
Second repeat  569   AAATTTATTGTTGG-TGTTA-CACTATATTATGTACATATGAAGAAACTTAGATATTCCA  626

First repeat   109   ACCTCGTGTTCAACATAAGTTTCAGAAAG------------------ATGGTACATATAT  150
Second repeat  627   ACCTTGTGTTCAATATAAGTTTCAGAAATTCAACAATAATATAGAATATGCTATATGAAT  686

First repeat   151   TTGAGAGGAAAAGTAAGGAAATACTTACATAGACCAAGTCTATAGTTAT--------TTA  202
Second repeat  687   TTGAGAGGAAAAGTGGGAAAATACTTACGTAGACGGGGTCtataattatctcatgtattt  746

First repeat   203   ATATTATCTTCATGAAATATAGTA-AATCATAA--ATAGTTTATAGACCAACTCTATATA  259
Second repeat  747   attttatttttataaaatataatataatcacaataatggtttatAGACCAGGTCTATATA  806

First repeat   260   AGTATTATCCGAAAAATTGATACATGCTTCAGGTGTGCACGTCGACATTTCAAAGTTCTA  319
Second repeat  807   AGTATTATCCGAAACAGTGATACATGCTTTAGGTGTGCGCG--GACATTT-AAAGTTCtt  863

First repeat   320   ATTTTTTGTT-----GGACAAACTTGAGTAATATTTCGTGTGTTATTGTAGAAATTGCTA  374
Second repeat  864   ttttttttttttGTGGGACAAACTTGTGTAATA---CGT-TGTTATTGTAGAATTTGATA  919

First repeat   375   TATAGGCAATGAATTTGAGACAAATAATTGTGTGATTTTATAATTCATACTCATTAATTA  434
Second repeat  920   TATGTGCAATGAATTTGTGACAAATAACTGTGTGATTTTACTATTCGTACTCATTATTTA  979

First repeat   435   -TACTTATATAACTTAATCTATATTTTAGGTAAATTGCATGTGGCATTGATCTACATTGA  493
Second repeat  980   CTACTTATATAACTTAATCTATATTTTAGGTAAATTGCATGTGCCATTGATCTACAATGA  1039

First repeat   494   CGAAAGTGACATCCTTTAACAATATTAGGTGTGGATCTAAATCTC---TAATAAAGATGA  550
Second repeat  1040  CGGATGTGACATCCTTTAACAATATTAGGTGTGGACCTAAATCTCTTATAACAAAGATGA  1099

First repeat   551   AATTGACATCATTTGACAAAATTTA--TTGTT  580
Second repeat  1100  AATTGACATCATTTGACAAAATTTAAGTTGTT  1131
```

Figure 14. The imperfect tandem repeat in the second intron of *Inc* The two repeated sequences from the *AmInc* allele were compared to each other. The first repeat was 527 bp and second repeat was 562 bp.

(a)




```
AcInc           CATTGCACTTTCCAAATCCTGCACCAGATGCAATAGTTTATCCATTGATAATCAGGGAAG 60
AmoInc          CATTGCACTTTCCAAATCCTGCACCCGATGCAATAGTTTATCCATTGATAATCAGGGAAG 60
AmInc           CATTGCACTTTCCAAATCCTGCACCAGATGCAATAGTTTATCCATTGAAAATCAGGGAAG 60

AcInc           GAGAGGAATCAATAATGGAAGAGCCAATCACTTTCACTGAAATGTACAAGAGGAAAATGG 120
AmoInc          GAGAGGAATCAATAATGGAAGAGCCCATCACTTTCACTGAAATGTACAAGAGGAAAATGG 120
AmInc           GAGAGGAATCAATAATGGAAGAGCCAATCACTTTCACTGAAATGTACAAGAGGAAAATGG 120

                   Early Stop codon
AcInc           GCAAGGATCTTGAGCTAGCTAAGCTAAAGAAATTGGCAAAAGAGAAAAAATTGCAAGAAG 180
AmoInc          GCTAGGATCTTGAGCTATCTGAGCCAAAGACATTGGCAAAAGAGAAAAAATTGCAAGAAG 180
AmInc           GCAAGGATCTTGAGCTAGCTAAGCTAAAGAAATTGGCAAAAGAGAAAAAATTGCAAGAAG 180

                  Stop codon
AcInc           AGGAGCTTGAGAAGTCCAAAATTGAAGTCAAAATAATATCCCGCAAAGGTATTGAAGAGA 240
AmoInc          AGGAGCTTGAGAAGTCCAAAATTGAAGTCAAAATAATATCCCGCAAAGGTATTGAAGAGA 240
AmInc           AGGAGCTTGAGAAGTCCAAAATTGAAGTCAAAATAATATCCCGCAAAGGTATTGAAGAGA 240

AcInc           TTCTAGCTTGATTGGTGAAAGGGAAACCGAAGAAAACGTGGATTTTTATTTTTATTTTTT 300
AmoInc          TTCTAGCTTGATTGGTGAAAGGGAAACCGAAGAAAACGTGGATTTTTATTTTTATTTTTT 300
AmInc           TTCTAGCTTGATTGGTGAAAGGGAAACCGAAGAAAACGTGGATTTTTATTTTTATCTTGT 300

AcInc           CTATTTTGGATTTTGTTGTAGTGTGAAACGATTTTTTATTAGCATCGATGAAGAAAAGGA 360
AmoInc          CTATTTTGGATTTTGTTGTAGTGTGAAACGATTTTTTATTAGCATCGATGAAGAAAAGGA 360
AmInc           CTGTTTTGGATTTTGTTGTAGTGTGTAACGATTTTTTATTAGCATCGATGAAGAAAAGGA 360

AcInc           AGACCATCTCAATAAATCTTCTTTGATGTAAGATGCGTTGTGGGAATTTGTCAAAAAAAA 420
AmoInc          AGACCATCTCAATAAATCTTCTTTGATGTAAGATGCGTTGTGGGAATTTGTCAAAAAAAA 420
AmInc           AGATCATCTCAATAAATTTTCTTTGATGTAAGATGCGTTGTGGGAATTTGTCAAAAAAAA 420

AcInc           AAAAA 425
AmoInc          AAAAA 425
AmInc           AAAAA 425
```

(b)

```
AmInc           LHFPNPAPDAIVYPLKIREGEESIMEEPITFTEMYKRKMGKDLELAKLKKLAKEKKLQEE 60
AcInc           LHFPNPAPDAIVYPLIIREGEESIMEEPITFTEMYKRKMGKDLELAKLKKLAKEKKLQEE 60
AmoInc          LHFPNPAPDAIVYPLIIREGEESIMEEPITFTEMYKRKMG*DLELSEPKTLAKEKKLQEE 60

AmInc           ELEKSKIEVKIISRKGIEEILA*
AcInc           ELEKSKIEVKIISRKGIEEILA*
AmoInc          ELEKSKIEVKIISRKGIEEILA*
```

Figure 15. Allelic variation in last exon of *Inc* creating the premature stop codon in *A. molle* (a) Last exon and 3'downstream region. Twelve SNPs were found between *AmInc* and *AmoInc* (position 26, 49, 86, 123, 138, 14, 151,291, 303, 326, 364 and 378) and six SNPs between *AmInc* and *AcInc* (position 49, 291, 303, 326, 364 and 378). (b) One SNP in *AmoInc* found to create the premature stop codon that cause the predicted peptide chain 42 amino acid shorter (* indicate stop codon).

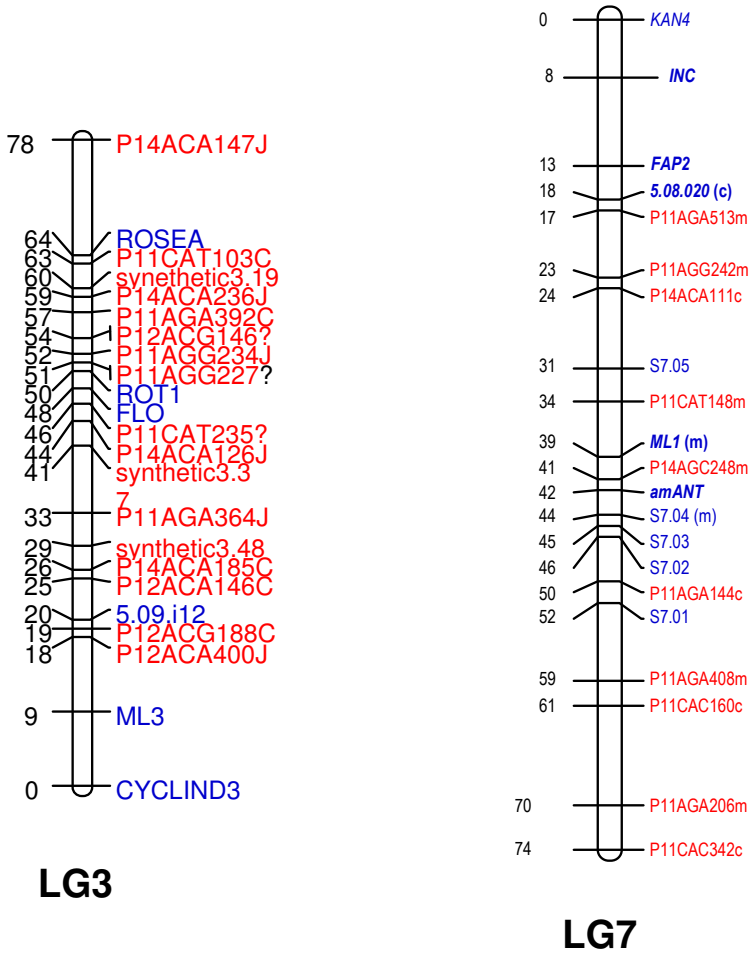


Figure 16. **Mapping *Ros* and *Inc* into the F_2 population linkage map** JoinMap 3.0 (Van Ooijen and Voorrips., 2001) was used to map *Ros* to LG3 around position 64 cM with a LOD score of 10. *Inc* was placed in LG7 around position 8 cM with LOD score of 7.

3.5 The correlation of candidate markers to phenotypic variation in an F_2 population

Two major QTL underlying differences in flower colour between *A. majus* and *A. charidemi* had been mapped to LG3 and LG7 in the proximity of *Ros* and *Inc*, respectively. Since these loci differed in structure between the two QTL mapping parents, their involvement was investigated further by testing the correlation between *Ros* and *Inc* genotypes and colour phenotypes.

Traits of interest, lobe and tube colour, from the F_2 population (V1640) were tested for their frequency distribution. Tube colour variation clearly showed a bell-shaped distribution (Figure 17b), suggesting that the trait was under the control of multiple genes. The distribution of lobe variation is more complicated and not likely bell-shaped (Figure 17a), suggesting that inheritance of the phenotypes involved epistasis or that segregation of genes affecting this trait in the F_2 is not fully independent. However, lobe and tube variation appeared to be affect by the same two major QTLs, as the traits are significantly correlated (P-value = 6.46×10^{-38}) (Figure 17c), The effects of the QTLs on each trait might therefore be different, which needed to investigate further.

Polymorphisms detected between *A. majus* and *A. charidemi* alleles were used to map *Ros* and *Inc* in the F_2 population used for QTL analysis. As expected, *Ros* mapped in LG3 and *Inc* in LG7 (Figure 16). Correlations between the genotypes of F_2 plants and their flower colour phenotypes were then examined. Data analysis showed correlation between *Ros1* genotype and flower tube color variation was strong, with P-value (from one-way *ANOVA*) at 3.6×10^{-22} and can explain about half of the flower tube colour variation (50%) in this F_2 population (Figure 20b) . This result also was supported by field observation by Whibley, which found that a pale tube flower was a good indicator for the *Ros1* allele from the pale species in hybrid zone (Whibley, 2004). *Ros1* and lobe flower colour also were found to show significant correlation (P-value = 1.86×10^{-10})

though the variance that *Ros1* phenotype can explain in population was lesser (26%) (Figure 20a).

In summary, a correlation was found between *Ros1* genotype and both traits; with a stronger correlation to tube colour than lobe colour (Figure 18, 19, 20a,b). The result suggested that plants homozygous for the *A. chardemi ros1* allele had paler tubes.

However, though the significant correlation of *Ros1* genotype to flower colour variation were shown by statistical analysis. An exception phenotype that cannot explain by *Ros1* alone was found in this population, i.e homozygotes *Ros1* from *A. majus* that gave pale tube or lobe (Figure 18, 19). This circumstance can be explained by two possible ways; 1) there is the recombination between the QTL and the candidate gene (*Ros1*) or 2) *Ros1* had been affected by other genes that segregating differently in F_2 population. The possible explanation was difficult to distinguish by F_2 population observation so the experiment were carried out in near isogenic lines (NILs) which are discussed later in Chapter 4.

Mapping of *Ros1* gene into the linkage map showed that *Ros1* was fitted into position 64 cM (Figure 16) with LOD score at 10. The most likely position of LG3 QTL was mapped at 59-60 cM, 4cM away from *Ros1* position. This would explain some exception occurred in F_2 population if there were some crossing over during meiosis to generate gametes in F_1.

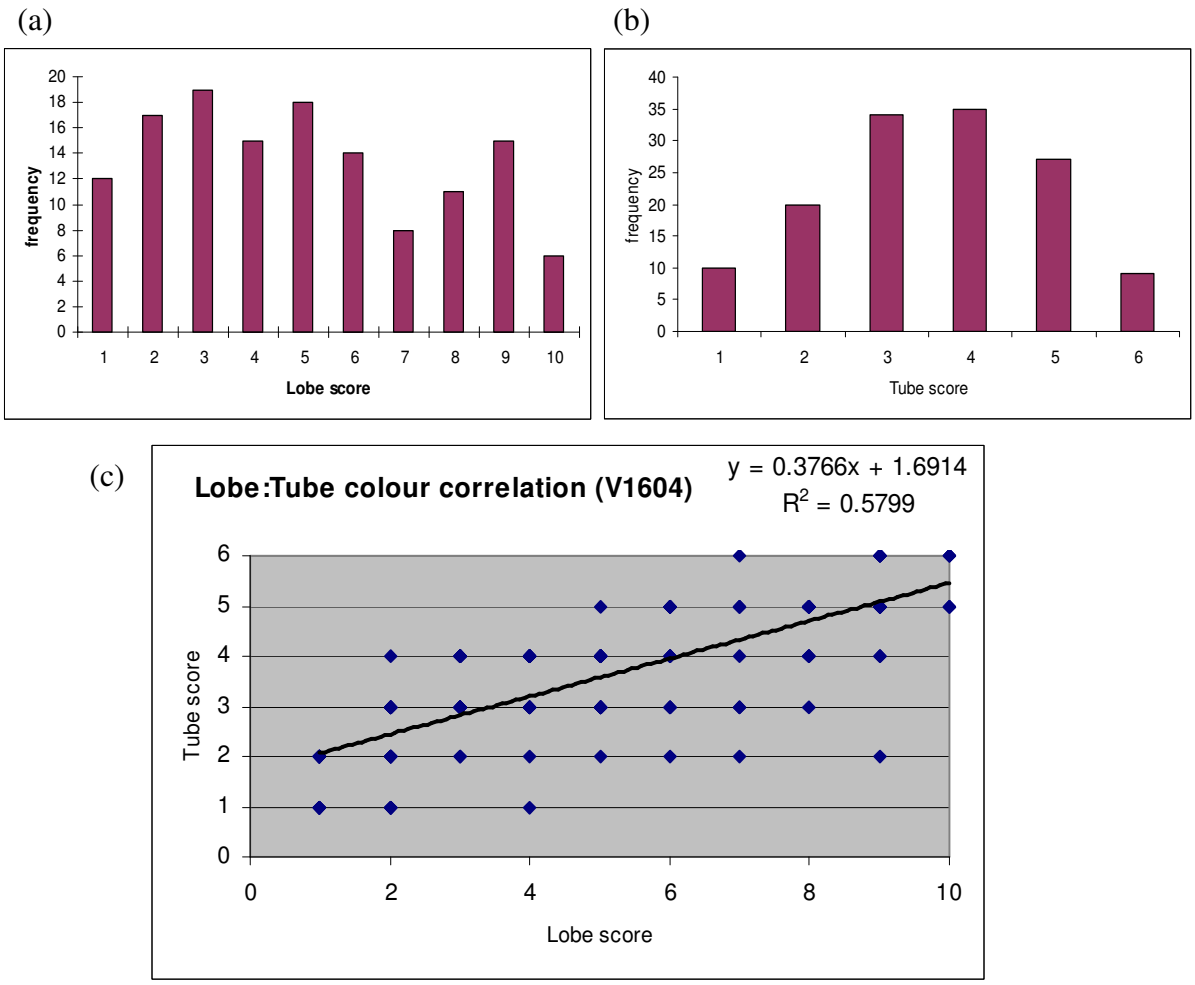


Figure 17. The phenotypic distribution and correlation in V1640 population (a) Flower lobe colour shows a complicated distribution in the F2 population. (b) Flower tube colour shows a bell-shaped distribution. (c) Though the distributions are different, the traits show significant correlation to each other at P-value = 6.46×10^{-38}.

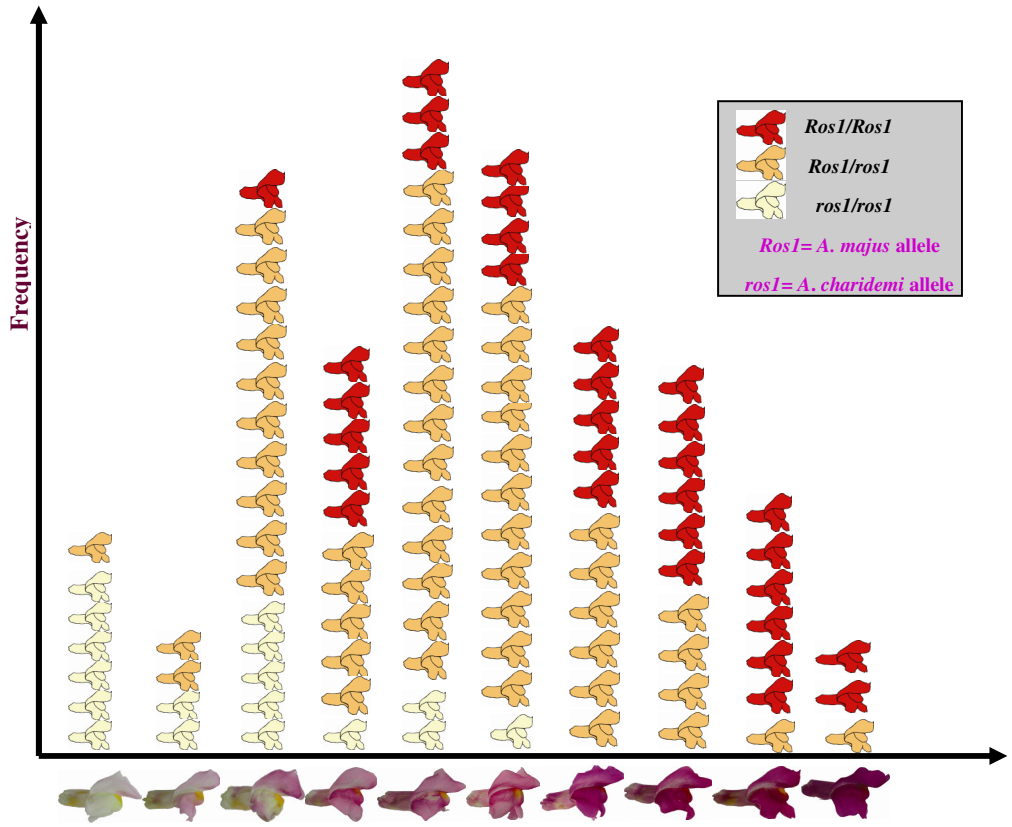


Figure 18. ***Ros1*** **genotype and petal lobe colour** The histogram shows the frequency of different lobe colour phenotypes in the F_2 mapping population (n=204). The *A. majus* parent was as darkly pigmented as the right-most flower on the x-axis and the *A. charidemi* as the third flower from the left. The *Ros1* genotypes of individuals are represented by the colour for the flower symbols (see insert). Plants homozygous for the *Ros1* allele from *A. charidemi* showed variation from very pale to intermediate colour, *A. majus* homozygotes showed variation from rather pale to very dark.

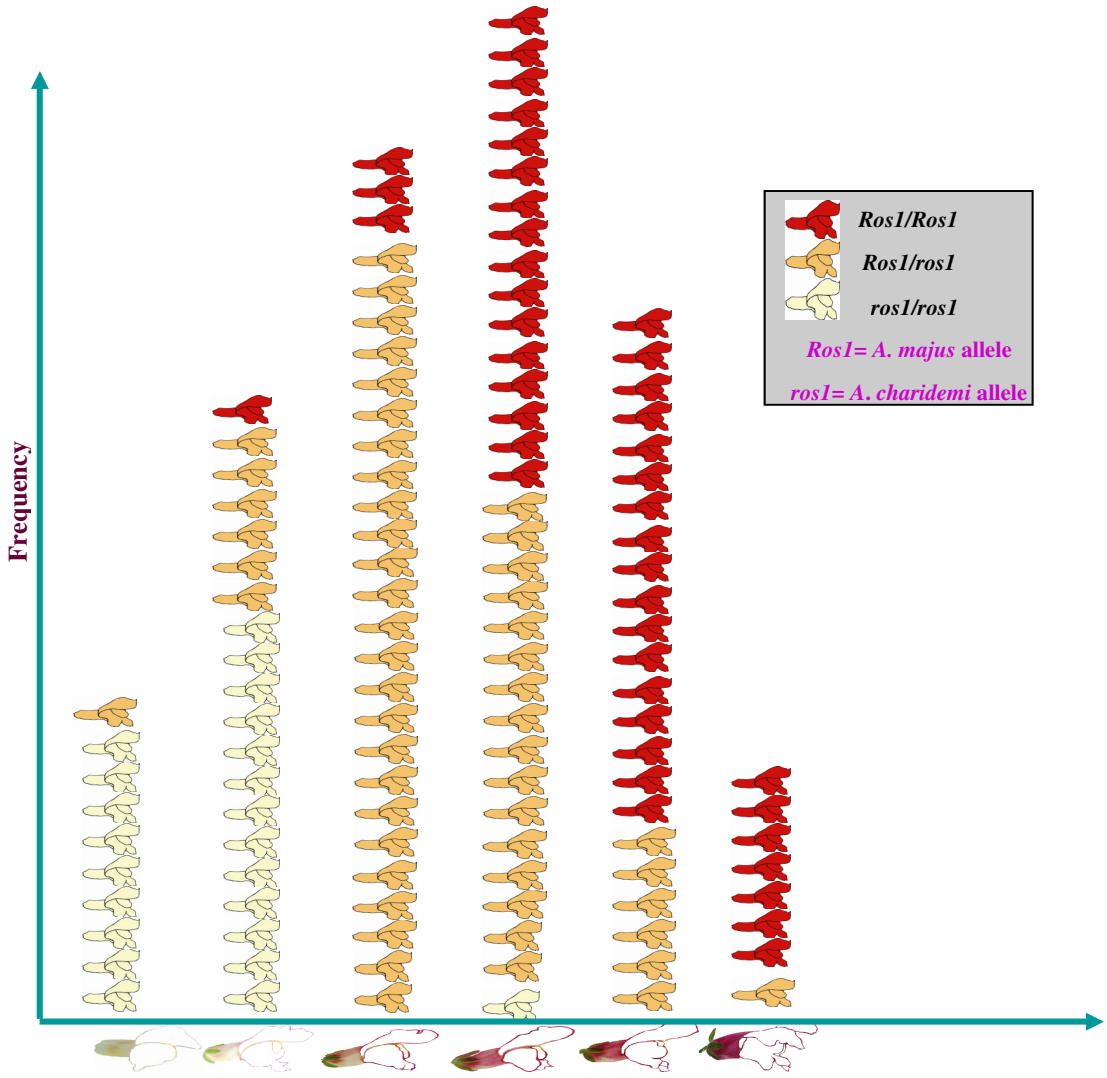


Figure 19. *Ros1* genotype and petal tube colour Tube colour variation showed bell shape distribution; though shift to be paler than the normal distribution. Homozygous ros1 from *A. charidemi* showed tendency to be pale with only one exception. Varying phenotype in heterozygous *ROS1* also found in this trait. *Ros1* homozygous from *A. majus* showed non-distinguished of colouring pattern. Ratio of *Ros1* genotype is 1:2:1.

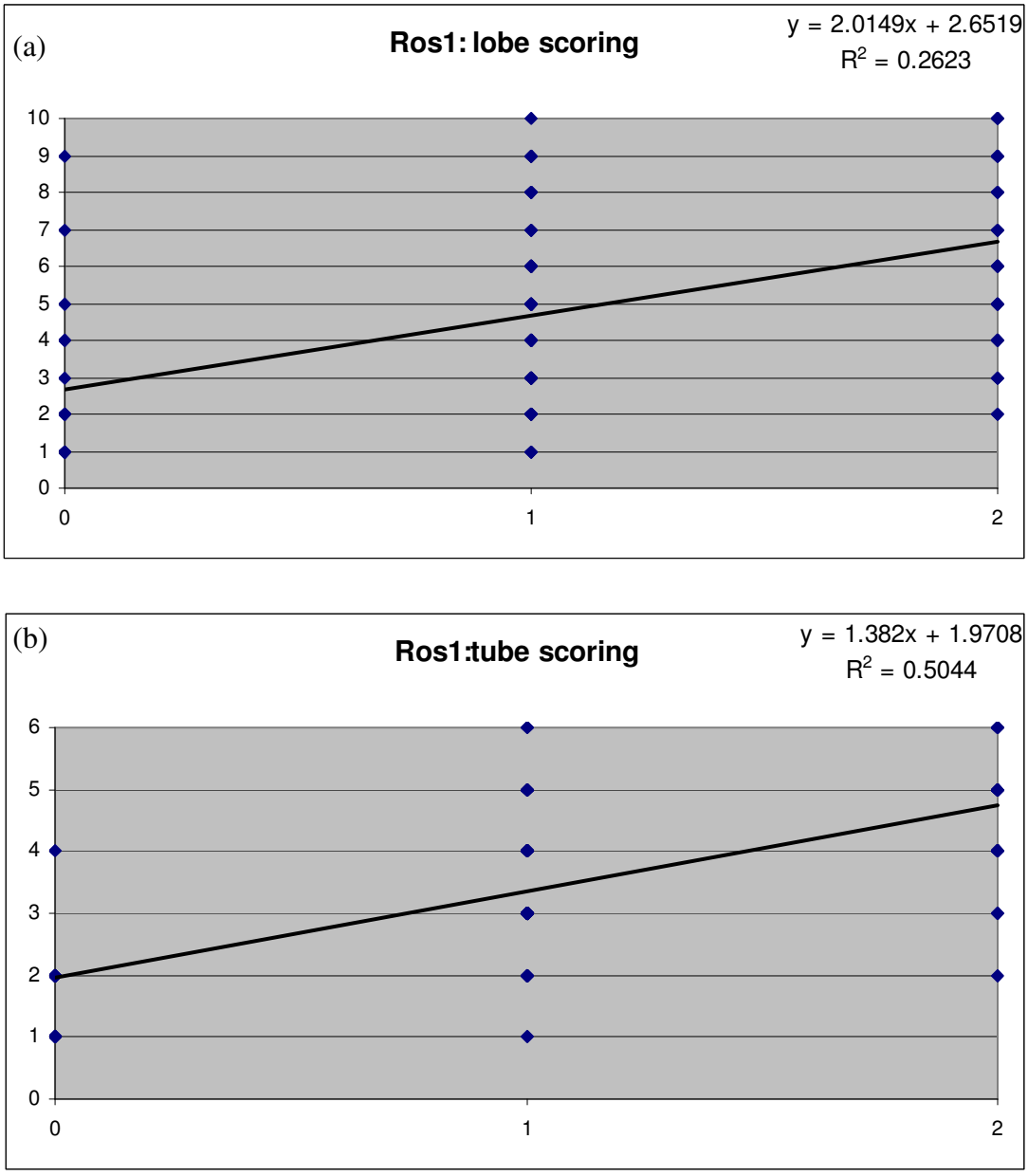


Figure 20. Simple linear regression of colour variation and *Ros1* genotypes in F_2 family V1640 (n = 135) The x-axis shows the dose of the *A. majus Inc* allele, so *A. charidemi* homozygous = 0, heterozygous = 1, *A. majus* (JI7) homozygous = 2, with lobe colour (a) or tube colour (b) each on a scale of 0-10 plotted against allele dose. (a) For lobes, R^2 = 0.2623 and the P-value for colour being correlated with genotype (from one-way *ANOVA*) 1.86 x 10^{-10}. (b) For tube colour, R^2=0.5044 and P=3.6 x 10^{-22}.

Lobe and tube colour variation showed less distinct patterns when comparing to *Inc* genotypes (Figure 21, 22), suggesting that *Inc* might have more subtle effect on flower colour variation. The regression analysis showed that *Inc* had significant correlation to lobe colour (P-value = 6.04×10^{-6}) and could explain about 14% of the lobe colour variation in the population (Figure 23a). *Inc* showed a marginal significant (P=0.01) correlation to tube colour variation, it could only explain 7% variance of tube colour (Figure 23b).

Inc genotypes showed significant correlations to both traits. However, mapping *Inc* put it at LG7 at position 8 cM (Figure 16), which very near one of the small QTLs (LG7, position 7 cM) affecting lobe variation, but 38 cM away from the second major-effect QTL (LG7, position 46 cM). This result raised the question whether *INC* was a candidate for the second major-effect QTL or only linked to it.

This also raised the question whether *Inc*, if it was not the second major-effect QTL, was the smaller QTL detected in LG7? The QTL at 7 cM appeared to show complete dominance, whereas the QTL linked to *INC* appeared to show semi-dominance (Figure 21, 22), suggesting that *INC* was not the minor-effect QTL However, it is possible that the effects of the minor-effect QTL cannot be separated from the effects of the linked major-effect QTL in the F2 population.

Nevertheless, *Inc* can be used as a marker for generating the new near isogenic lines (NIL) families in attempts to identify chromosome regions that are responsible for traits of interest and to separate the effects of linked QTL. Therefore linkage of *INC* to flower colour QTL observed in F_2 was tested further in NILs (Chapter 4).

Another interesting observation in the F_2 population was that the palest flower was paler than the *A. charidemi* parent (Figure 18, 21), which is also discussed further in Chapter 4.

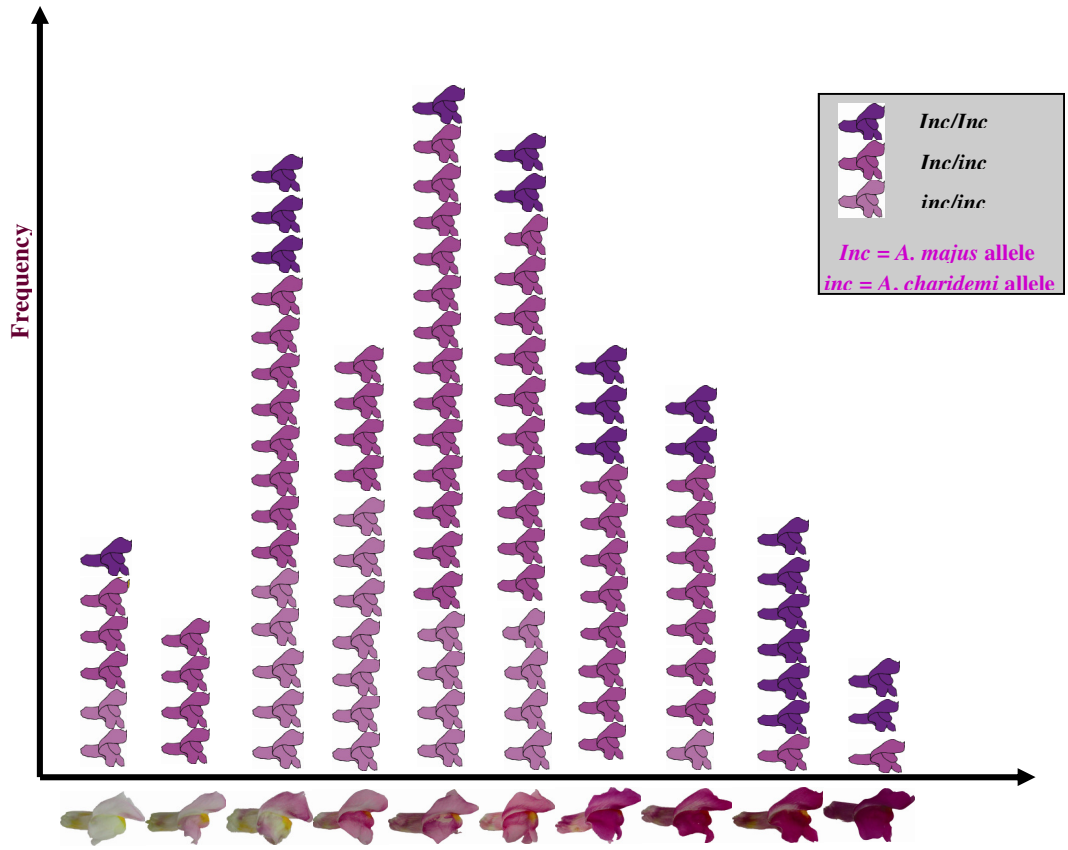


Figure 21. *Inc* genotype and petal lobe colour Pictograms show correlation between *Inc* genotype and phenotypic variation in F_2 population (n=204). Y-axis is the frequency of the variation with each flower represents individual. X-axis is the phenotypic variation represents by real picture to reflect the scoring. *A .charedemi* parent flower colour resembled the third flower of Y-axis from left. *A. majus* parent flower colour resembled the tenth flower. (a) Lobe colour variation in F_2 population shows normal distribution suggesting semi-dominance effect. Genotypic variation of *Inc* showed less correlated pattern. Ratio of *Inc* genotype is 1:2:1.

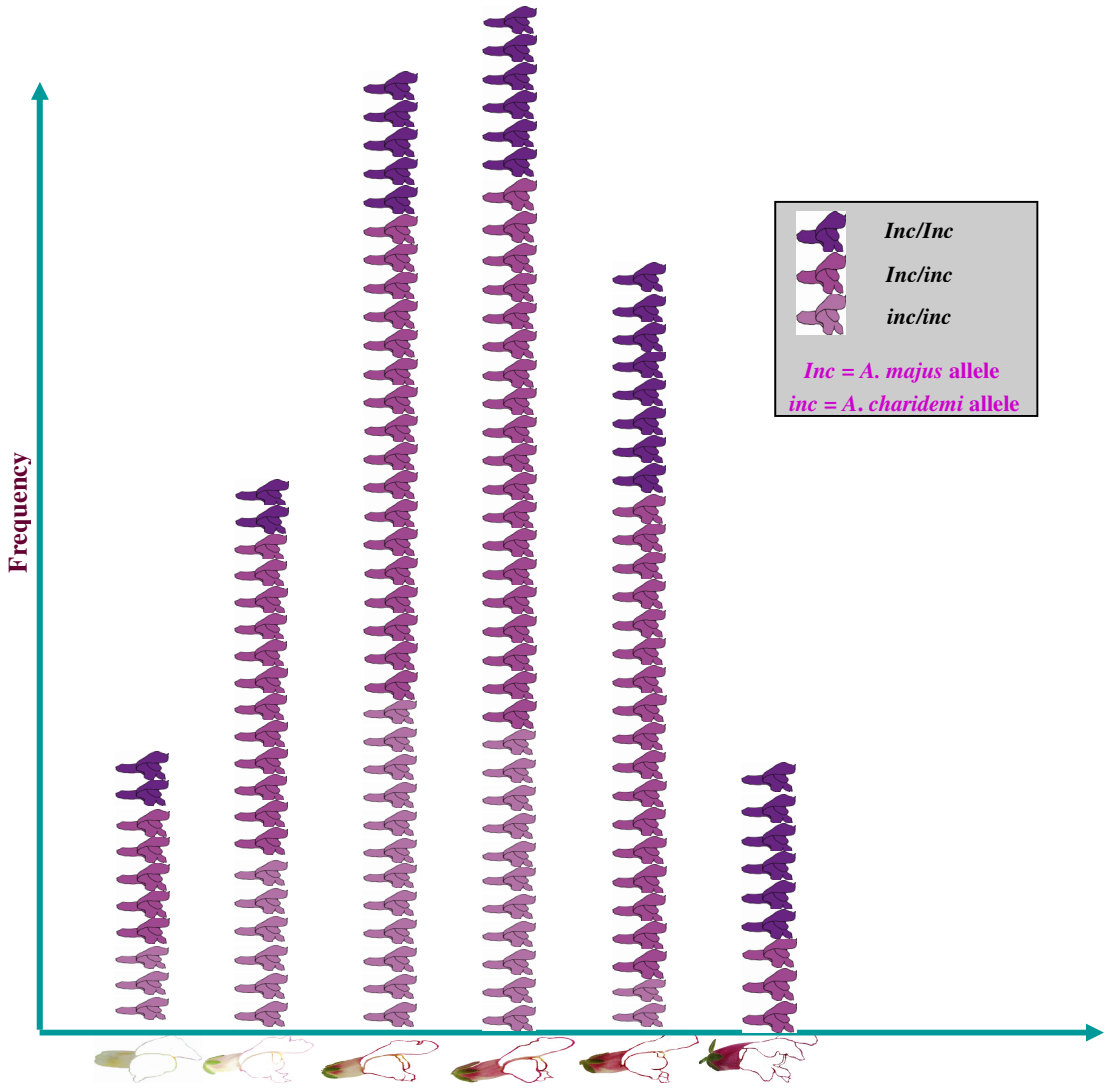
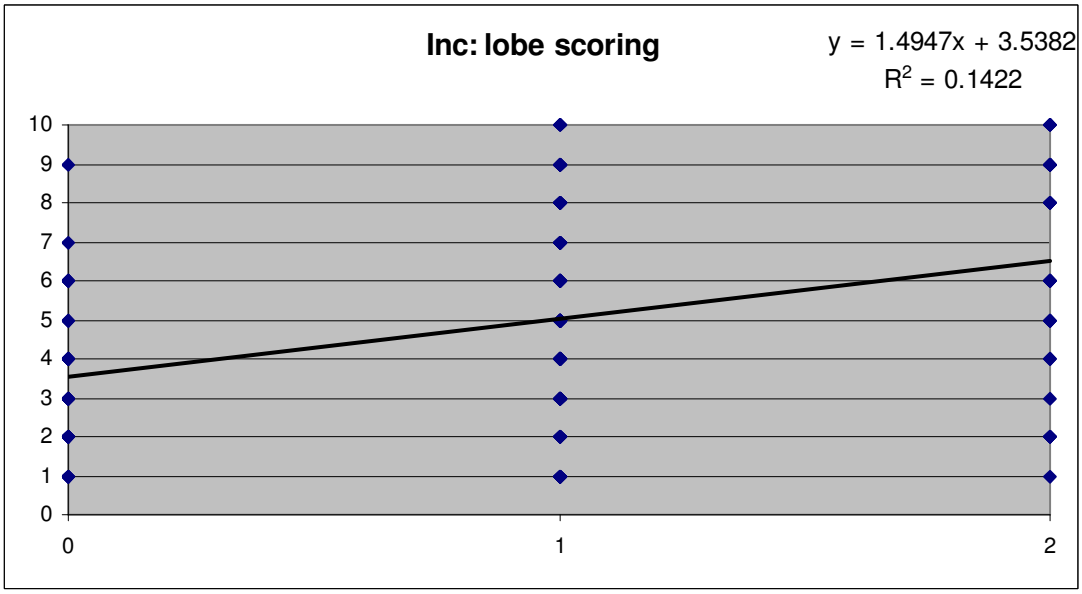


Figure 22. *Inc* genotype and petal tube colour Tube colour variation showed bell shape distribution. INC genotypes also showed less-correlated pattern towards tube colour variation. Ratio of *Inc* genotype is 1:2:1.



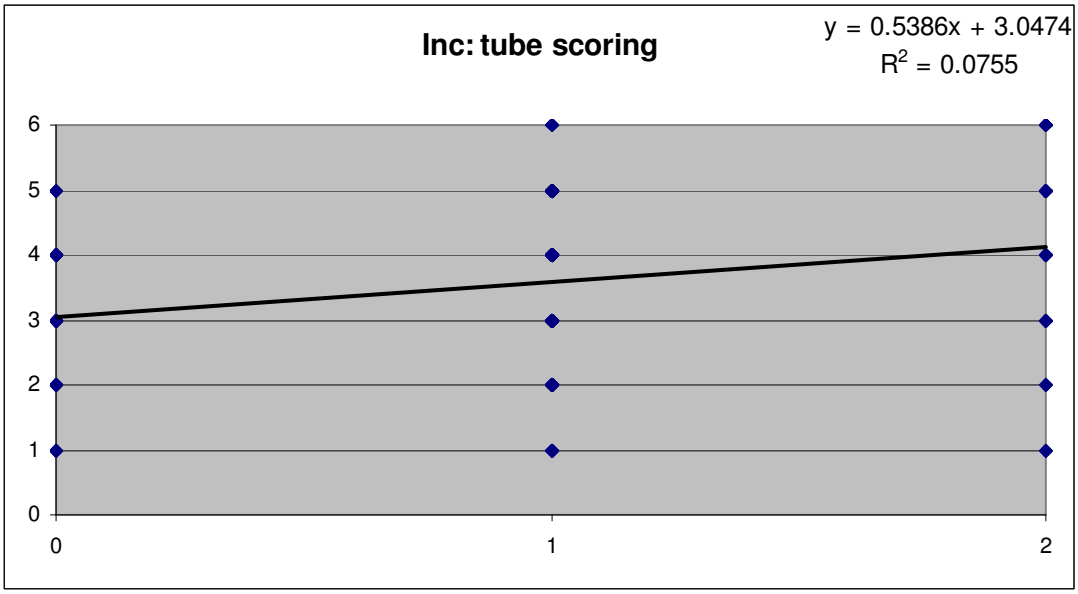


Figure 23. Simple linear regression of colour variation and *Inc* genotypes in F_2 family V1640 (n = 135). The x-axis shows the dose of the *A. majus Inc* allele, so *A. charidemi* homozygous = 0, heterozygous = 1, *A. majus* (JI7) homozygous = 2, with lobe colour (a) or tube colour (b) each on a scale of 0-10 plotted against allele dose. (a) For lobes, R^2 = 0.1422 and the P-value for colour being correlated with genotype (from one-way *ANOVA*) 6.04 x 10^{-6}. (b) For tube colour, R^2=0.0755 and P=0.01.

3.6 Conclusion and Discussion

In this chapter, QTL analysis in an *A. majus* x *A. charidemi* hybrid suggested that flower colour variation is genetically complex. Some QTLs were found to be specific to either lobes or tubes. This is consistent with the idea that different mechanisms control anthocyanin production in different parts of the flower. However, most of the variation in both tube and lobe colour could be explained by two major effect QTL.

Some F_2 plants were paler than their *A. charidemi* parent. A similar phenomenon had been found when the same inbred line of *A. majus* (JI7) had been crossed with *A. molle* (Whibley, 2004). This could be explained if the *A. majus* parent carries alleles that reduce anthocyanin pigmentation or if *A. charidemi* contributes negative dominant alleles that can make heterozygotes paler than *A. charidemi* homozygotes. However all alleles detected in QTL analysis acted in the parental direction (i.e, the *A. majus* allele increased colour relative to the *A. charidemi* allele) and none showed complete dominance.

The two major effect QTL mapped in LG3 and LG7 as did two candidate genes – *Ros* (LG3) and *Inc* (LG7). Both genes differed in structure between *A. majus* and *A. charidemi* as expected of QTL genes. *Ros1* and *Inc* were genotyped in members of the F_2 QTL mapping population, allowing direct genotype-phenotype correlations to be tested directly. This revealed an association between *Ros1* genotype and flower colour, particularly flower tube colour, with *A. charidemi* homozygotes tending to have paler flowers. However, ranges of flower colours were associated with each genotype. This can be explained in two ways. 1) *Ros* is not the QTL, so that the exceptions carry recombinations between the candidate gene and the QTL. 2) *Ros* is the QTL, but its effects are modified by alleles of other genes segregating in the F_2 population (epistasis), this aspect was supported by the extra WRKY recognition site that found only in

AmRos1 promoter region. However, more intensive molecular study such as promoter deletion analysis is needed to support this idea.

To distinguish between these possibilities, subsequent mapping was carried out in near isogenic lines (NILs) in which only part of the genome segregates (Chapter 4).

A likely loss-of-function *Ros1* allele was found in *A. molle*. In a small F_2 population, this co-segregated with white flower colour, suggesting that the allele might contribute to the pale flowers of this species. However, the effect of the mutation on flower colour was not tested directly and the possibility that other loci linked to *Ros1* were responsible could not be ruled out. A QTL for flower colour had previously been detected in LG3 at position 8 cM in *A. majus* x *A. molle* hybrids (Yang, 2007), which can possibly be *Ros1* though it was not certain. Scwhinn *et al*. (2006) also suggested that *Ros1* might underlie *A. molle* the pale flowers as she found no expression of *Ros1* in this species.

Chapter 4

Testing candidate genes in NILs families

In the previous chapter, two major QTLs were found to explain more than half of the variation in flower colour in the *A. majus* x *A. charidemi* F_2 population, V1640. Variation in petal lobe and tube colour in the V1640 population was tested for association with two candidate genes, *Ros1* and *Inc*, which had been mapped close to the locations of the two major QTL. A correlation could be found between the *Ros1* genotype and flower colour, particularly petal tubes. However, there was considerable variation in flower colour within each *Ros1* genotype class, suggesting either that *Ros1* was not one of the major QTL or that its effect was dependent on other genes. A correlation was also found between flower colour and *Inc* genotypes.

In this chapter, near isogenic lines (NILs) were used to test the involvement of *Inc* and *Ros1* further. The NILs had been produced by back-crossing an *A. majus* x *A. charidemi* hybrid to *A. majus* for four generations and therefore carried regions of the *A. charidemi* genome in the genetic background of *A. majus*. NILs heterozygous for regions LG3 and LG7 carrying the *Ros1* and *Inc* loci were identified and used to test whether these genes co-segregated with flower colour. The NILs were then used to test association between flower colour and other candidate genes, beginning with ABP genes known to map to LG3 or LG7. Other genes known to be involved in pigmentation were then tested for their association with flower colour variation, either by mapping in V1640 or NILs families and by expression analysis in NILs segregating for flower colour variation.

4.1 NILs families colour flower variation.

The NILs families were generated in the *A. majus* (JI.7) genome background (Figure 24). Each family remained heterozygous for a fraction of the genome. Families were genotyped at *Ros1* and *Inc* to identify a NIL that was heterozygous for both loci, in LG3 and LG7 respectively. It was self pollinated, so that the heterozygous regions segregated in its progeny (family M78) and four members of M78 were self pollinated to generate

four further families of progeny (Figure 24). Family M239 (n=17) was homozygous for *AcRos1* and *AcInc*; M241 (n=109) segregating *Ros1* but was homozygous for *AcInc*; M243 (n=22) was homozygous for *AmRos1* and for *AmInc* and family M244 (n=89) segregated at both *Ros1* and *Inc*. Flower colours were scored in these families in the same way as the F_2 population.

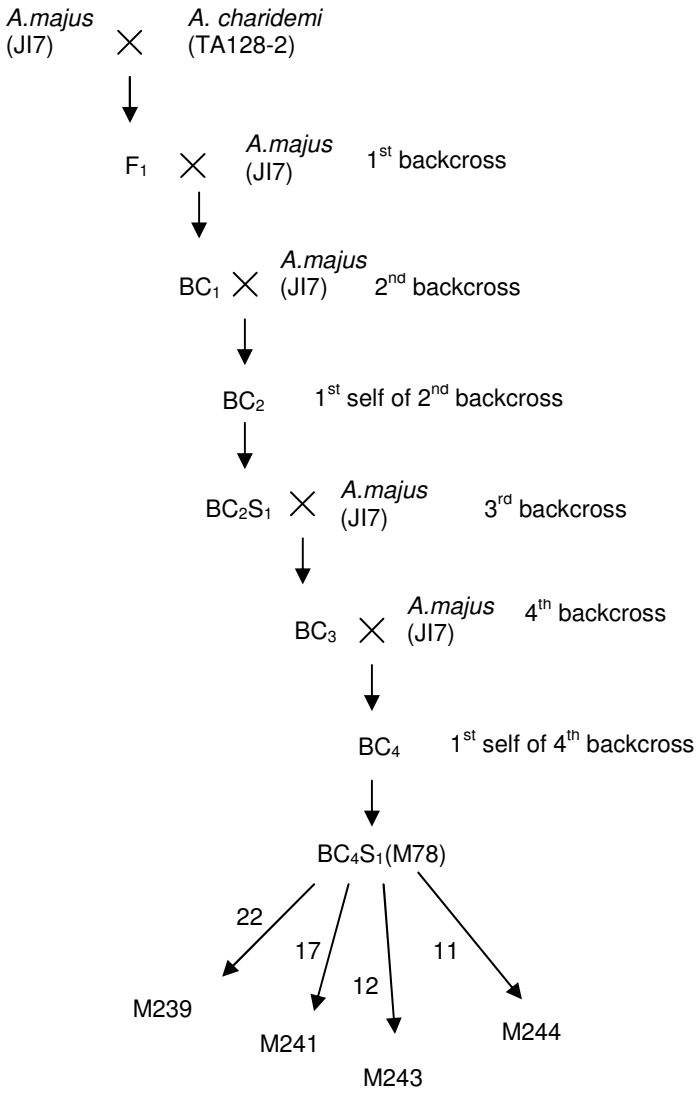


Figure 24. Pedigrees of NILs families NILs were generated by backcrossing an F1 hybrid four times to its *A. majus* parent (JI.7) and then selfed so that heterozygous regions segregated in their progeny (1st self of 4th backcross). Crosses and the first self pollination were carried out by Lucy Copsey at John Innes Centre. Family M78 was grown by Manuela Costa and genotyped for *Ros1* and *Inc* in Edinburgh in 2007. Its members were selfed by Joanna Cricthley in Edinburgh in 2009. The first self of the second backcross was carried out to screen for phenotypic differences between lines and so identify lines that carried *A. charidemi* alleles of genes of interest.

Family M243 did not show any segregation for flower colour; all individuals had dark flowers scored as 8 - 10 (Fig, 27a,c). This family was homozygous for *AmRos1* and *AmInc*, and therefore lack of variation for flower colour does not rule out involvement of either of these genes. Families M241, though generated to segregate only in LG3, segregated for petal lobe with a bell shaped distribution (Figure 26c), suggesting more than one single gene contribute flower lobe variation. On the other hand, the lobe colour segregation in M244, segregating for both *Ros1* and *Inc*, showed the similar pattern of approximate 3:1 ratio (69:16) (Figure 28c) to lobe colour segregation in M239 (Figure 25c), suggesting the stronger influence of dominant dark allele, possibly from LG7.

The tube colour distribution for M241 and M244 also were not bell-shaped distribution as found in V1640 (Chapter3; Figure 17b) but showed approximately 1(pale):1(dark) ratio (62:49) in M241 (Figure 26d) and 1(pale): 2(dark) (33:56) in M244 (Figure 28d) suggesting unknown genes affecting (possibly epistasis) *Ros1* might not be segregated in these families.

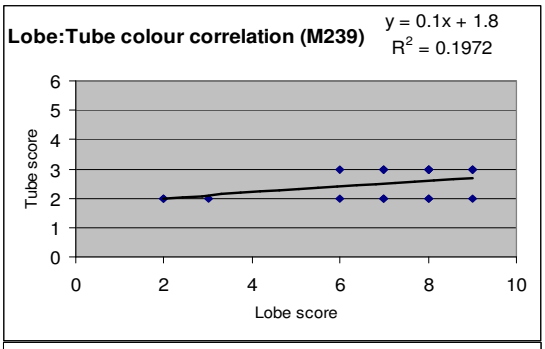
The significant correlation of lobe and tube phenotype in M241 and M244 were found (M241; $R^2 = 0.3246$, P-value = 6.74×10^{-11} [Figure 26b], M244; $R^2 = 0.4143$, P-value = 1.34×10^{-11} [Figure 28b]), however, the correlation was less compared to V1640 F_2 family ($R^2 = 0.5659$, P-value = 6.46×10^{-38}). Families M241 and M244 segregated at *Ros*, supporting the involvement of *Ros* or a gene closely linked to it, in colour variation. However, there was no correlation of flower lobe and tube colour in M239 ($R^2 = 0.1972$, P-value = 0.074 [Figure 25b]) which was homozygous for *AcRos1* and *AcInc*, suggesting that QTLs that were closely linked to *Ros1* and *Inc* or the genes themselves can contributes flower variation independently in tubes and lobes. The *Ros* and *Inc* genotypes from NIL families also suggested that *Ros1* contributes mainly to variation in tube colour.

M241 and M244 families showed a similar range of flower colour phenotypes, but only M244 segregated at *Inc*. This suggests that *Inc* did not contribute to flower colour variation in these families. This idea was further supported by family M239, which was homozygous for *A. charidemi* alleles at both *Ros1* and *Inc*, but segregated for petal lobe colour, suggesting involvement of a major effect locus that was not *Ros1* or *Inc*. Dark and pale flowered plants in M239 family occurred in a ratio of ~3 dark:1 pale (14:3) suggesting involvement of a single locus with a dominant dark allele, likely from *A. majus* (Figure 25c). This was consistent with the results of QTL analysis in the F_2 population, which estimated that the QTL at 7 cM in LG7 had a completely dominant *A. majus* allele that increased pigmentation.

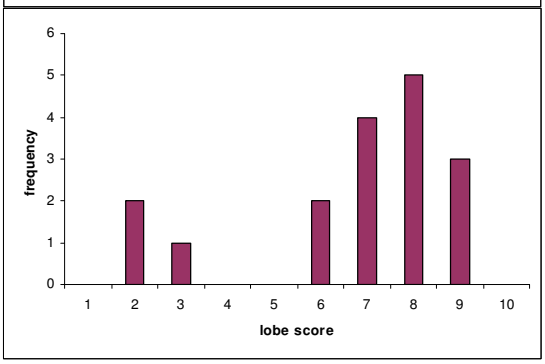
These findings suggest that *Ros* might contribute to differences in petal tube colour, although involvement of a linked gene cannot be ruled out. This is consistent with the structural differences in *Ros1* promoter found in *A. majus* and *A. charidemi* (Chapter 3). They also suggest that *Inc* is unlikely to be the second major-effect QTL, in LG7, but suggest that this major-effect QTL might segregate in some of the NILs families. These hypotheses were tested further, by genotyping NILs families for *Ros1* and *Inc*, as described below. The NILs were also used to test involvement of other candidate genes.

(a)

(b)



(c)



(d)

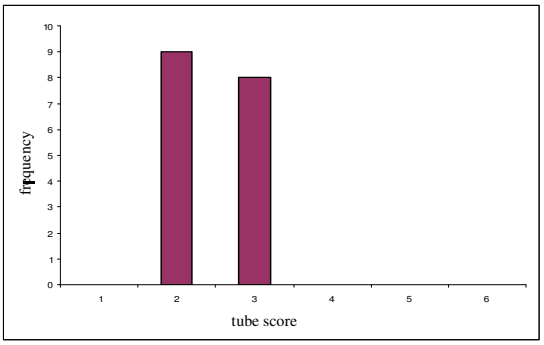


Figure 25. M239 NIL family was homozygous with *AcRos1* and *AcInc* (a) but still showed flowers with a range of colours. (b) Flower lobe and tube showed no correlation with each other (P= 0.074). (c) The segregation of petal colour fell into the estimated ratio 3 (dark): 1 (pale) (14:3) suggesting the involvement of a single locus with dominant dark allele. (d) The M239 family was homozygous for the *AcRos1* allele and showed pale tube colour (score 2-3), suggesting that *Ros1*, or a gene closely linked to it, is associated with tube colour variation.

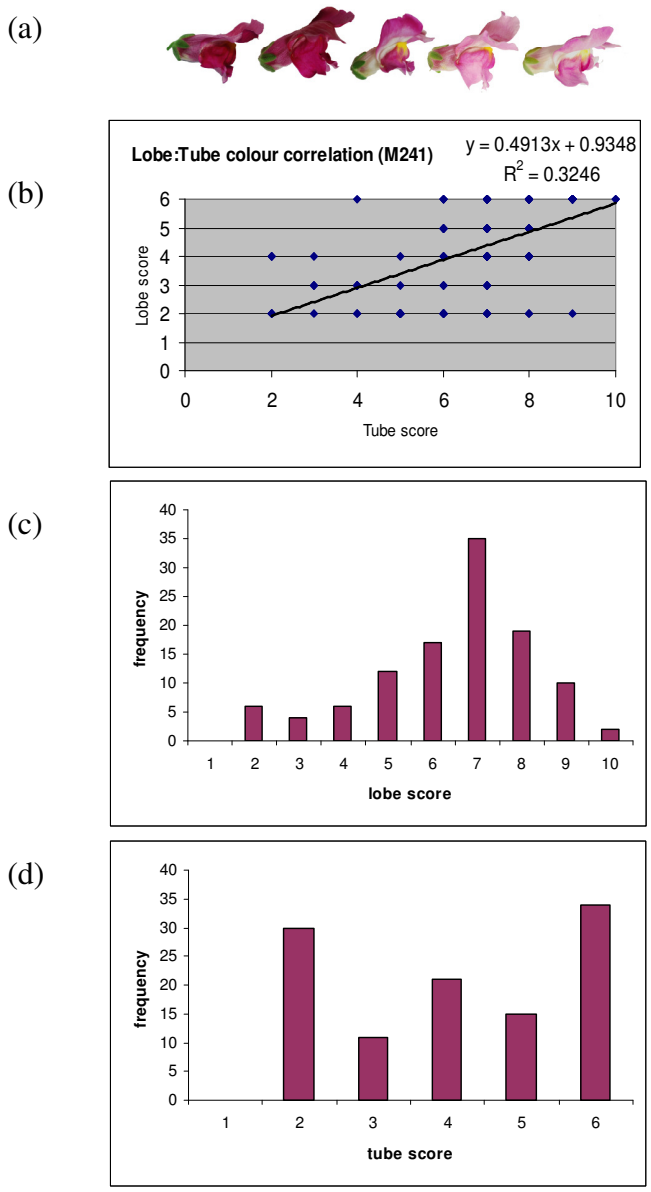
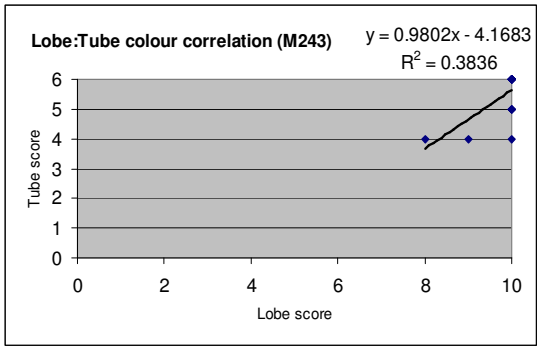


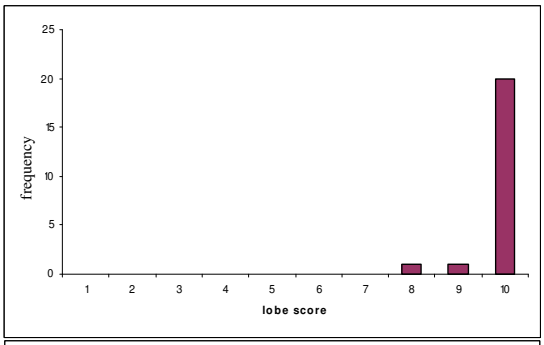
Figure 26. M241 NIL family was segregating for *Ros1* (a) Family showed flowers colour range in family. (b) Flower lobe and tube showed significant correlation (R^2 = 0.3246, P value = 6.74 x10^{-11}). (c) The segregation of petal colour is bell shaped distribution. (d) Tube colour were segregating for 1:1 ratio.

(a)

(b)



(c)



(d)

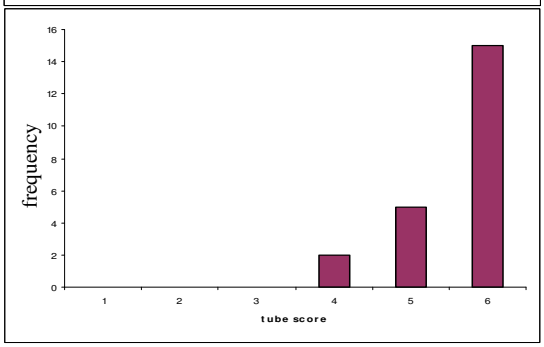
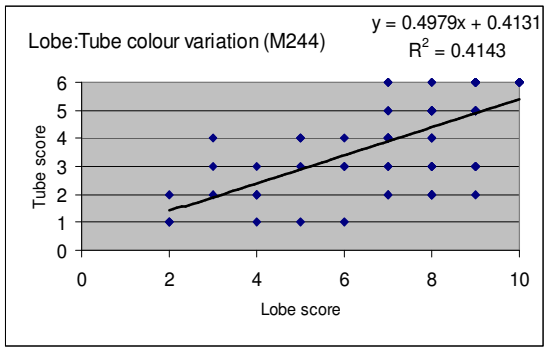


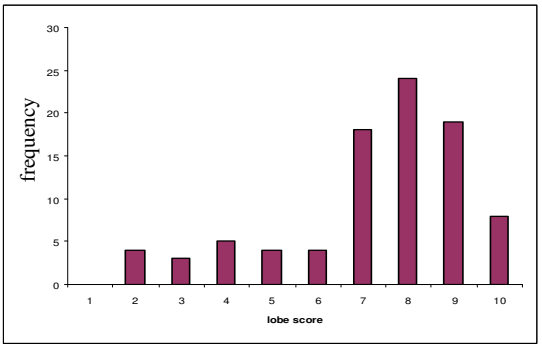
Figure 27. M243 NIL family was homozygous with *AmRos1* and *AmInc* (a) Family showed flowers colour range in family. (b) Flower lobe and tube showed significant correlation (R^2 = 0.3836, P value = 0.002). (c) Petal colour showed no segregation. (d) All tube colour were dark (score 4-6) when homozygous *AmRos1* suggesting the association of *Ros1* or a gene closely linked to it with tube colour variation.

(a)

(b)



(c)



(d)

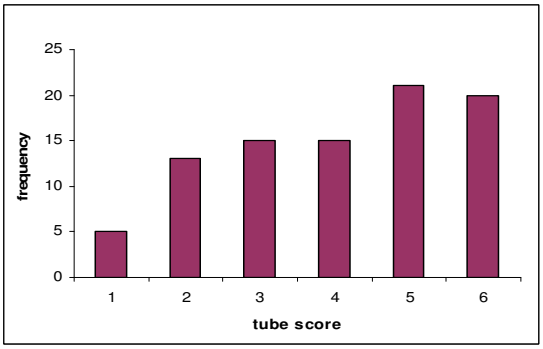


Figure 28. M244 NIL family was segregating for *Ros1* and *Inc* (a) Family showed flowers colour range in family. (b) Flower lobe and tube showed significant correlation (R^2 = 0.4143, P value = 1.34 x 10^{-11}). (c) The segregation of petal colour is similar to M239 with ration approximate 3(dark):1(pale). (d) Tube colour variation ratio of dark:pale was about 1:2.

4.2 Testing genes involved in Anthocyanin Biosynthesis Pathway (ABP)

Genes involved in the ABP can be divided into two groups, structural genes and regulatory genes. The structural genes encode enzymes required for anthocyanin biosynthesis (Figure 29). This group itself can be divided into two sub-groups, early biosynthetic genes (EBGs) which are required for anthocyanin and early flavonoid product production and late biosynthetic genes (LBGs) which are required only in anthocyanin production. The two groups show partly independent activation in dicotyledonous species (Martin and Gerats, 1993; Quattrocchio *et al*., 2006). Regulatory genes have been divided into three classes, encoding bHLH, MYB or WD proteins that function together in transcription factor complexes.

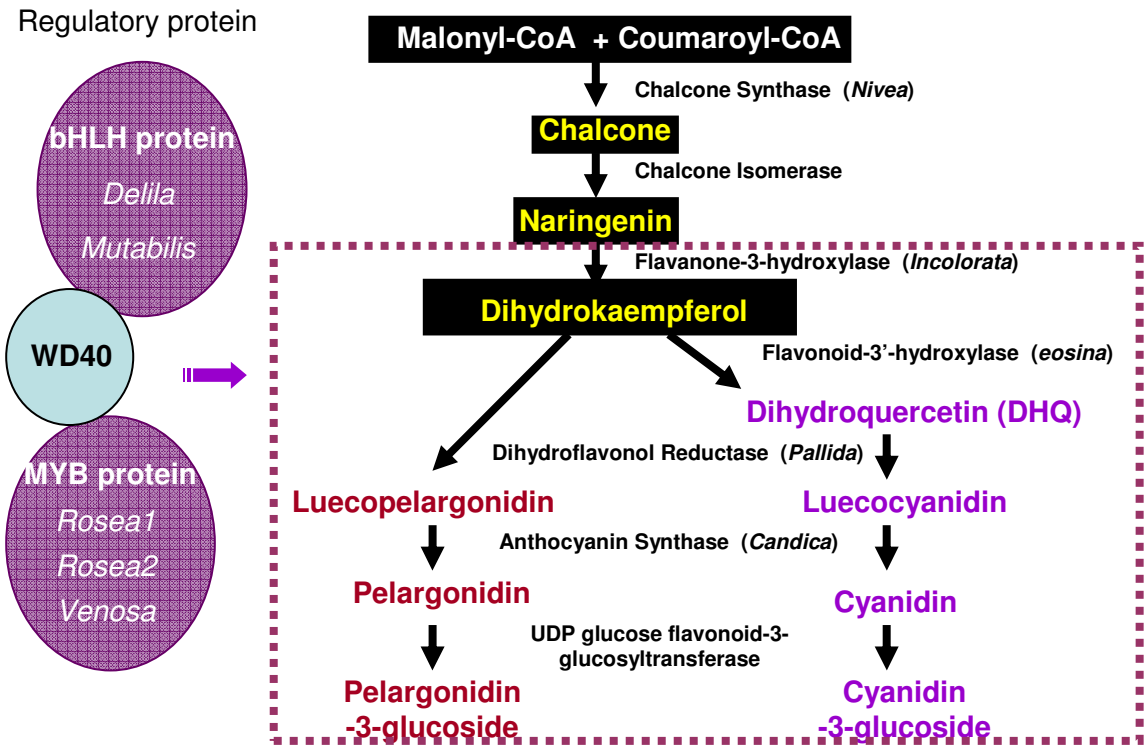


Figure 29. The anthocyanin biosynthesis pathway (ABP) in *antirrhinum* LBGs are shown in the dotted square and are regulated by transcription factor complexes containing bHLH-WD40-MYB proteins. *Antirrhinum* genes encoding ABP enzymes and regulators are shown in italics.

In an attempted to find candidate genes responsible for flower colour variation between *A. majus* and *A. charidemi*, both the structural and regulatory genes were considered, as discussed in further detail below.

4.2.1 Structural genes

i. *Nivea* (*Niv*)

Nivea (*Niv*), the first gene unique to the *Antirrhinum* ABP encodes chalcone synthase (CHS), which catalyzes the condensation of three molecules of malonyl-CoA with one molecule of *ρ*-coumaroyl-CoA to yield the yellow-coloured terahydroxychalcone (Holton and Cornish, 1995). *Niv* was originally reported to be a single copy gene in *Antirrhinum* (Holton and Cornish, 1995) although a second cDNA sequence was later identified (Hatayama *et al*., 2006). A third *Niv*-like gene can be detected in the recently available *A. majus* genome sequence (data not shown). *niv* loss-of-function mutations give albino flowers (Fig. 30, 31a). Other, semi-dominant *Niv* mutations have also been described that reduce pigmentation when heterozygous with a wild-type allele (Figure 30). One possibility is that the mutant alleles produce antisense RNA which inhibits expression of the wild-type allele, though no antisense RNA was detected (Bollmann, Carpenter and Coen, 1991). More recent understanding of epigenetics suggests that they effect might cause by siRNA though this was not tested at the time. The *Niv* gene had been mapped as a dominant markers in F_2 family V1640 by Feng *et al*. (2009) to 85 cM in LG1 (Figure 32a).

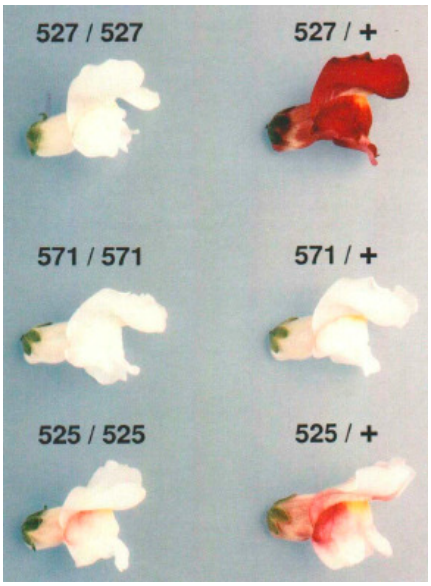


Figure 30. Allelic variation of *Niv* and its flower phenotypes A *niv-527* allele is fully recessive, homozygous of *niv-527* showed a white flower whereas heterozygous with wild-type allele gave a fully-pigmented flower. A *niv -571* allele is semi-dominant, though the homozygous of *niv-571* gave a white flowers, heterozygous with wild-type gave a pale flowers. A *niv- 525* allele is also semi-dominant, though the effect is not as strong as *niv-571* in heterozygous.

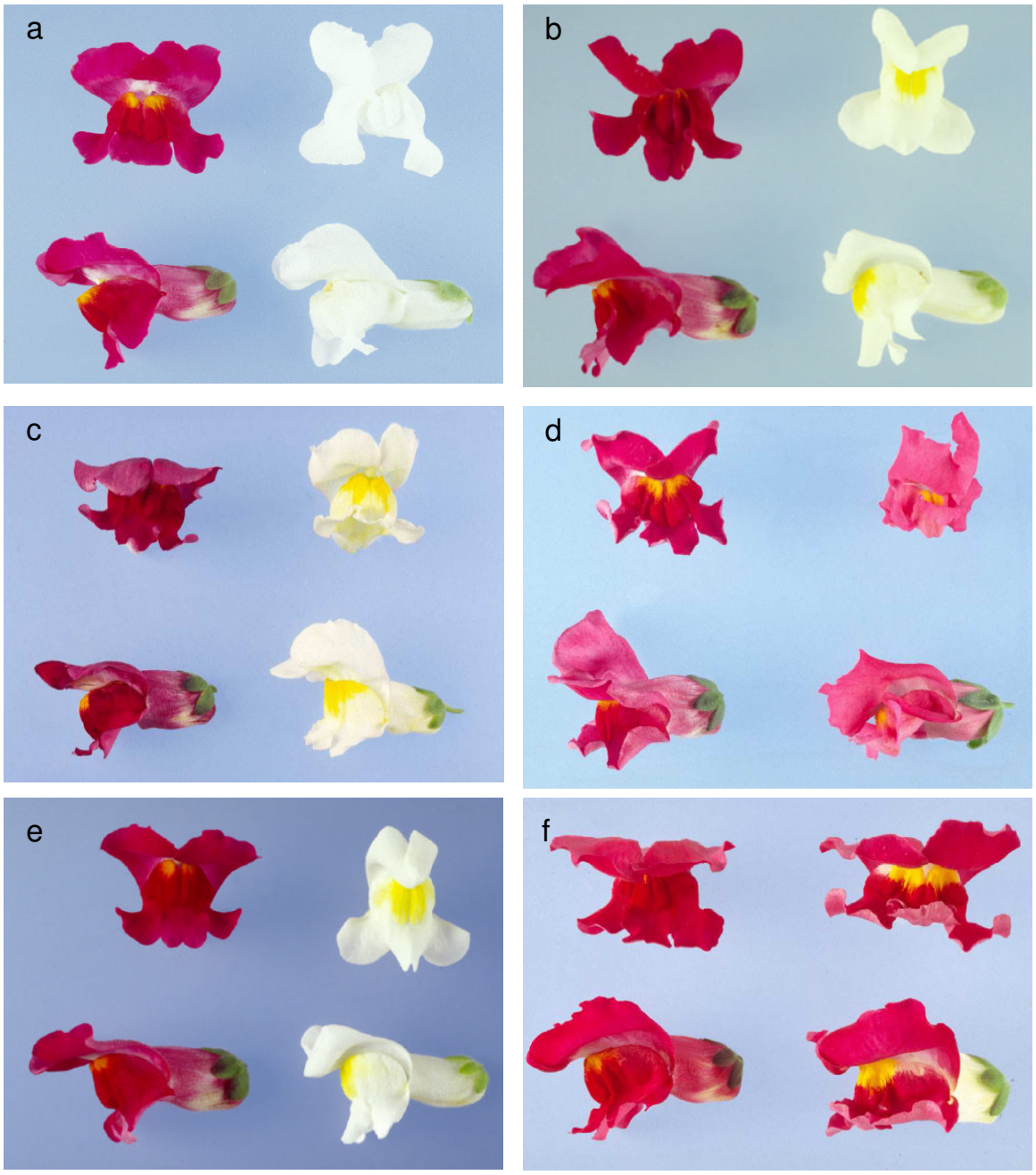


Figure 31. Effects of structural and regulatory mutants in the ABP Each panel shows mutant flowers to the right with wild-type flowers in the same genetic background to the left. (a) Loss-of-function *niv* mutations produce white flowers. (b) *inc* loss-of-function mutations gives ivory flowers with yellow face. (c) *pal* loss-of-function mutants produces flowers similar to *inc* mutants. (d) Loss-of-function in *eos* gives pinker flowers. (e) Loss-of-function in *candi* produces ivory flower with yellow faces, similar to *inc* and *pal* mutants. (f) Loss-of-function of *del* allows fully pigmented corolla lobes (where the bHLH regulatory function is provided by *Mut*) but ivory corolla tubes (where *Mut* is not expressed). All photos are from the JIC *Antirrhinum* stock collection, by Rosemary Carpenter with pictures taken by Andrew Davis and Peter Scott.

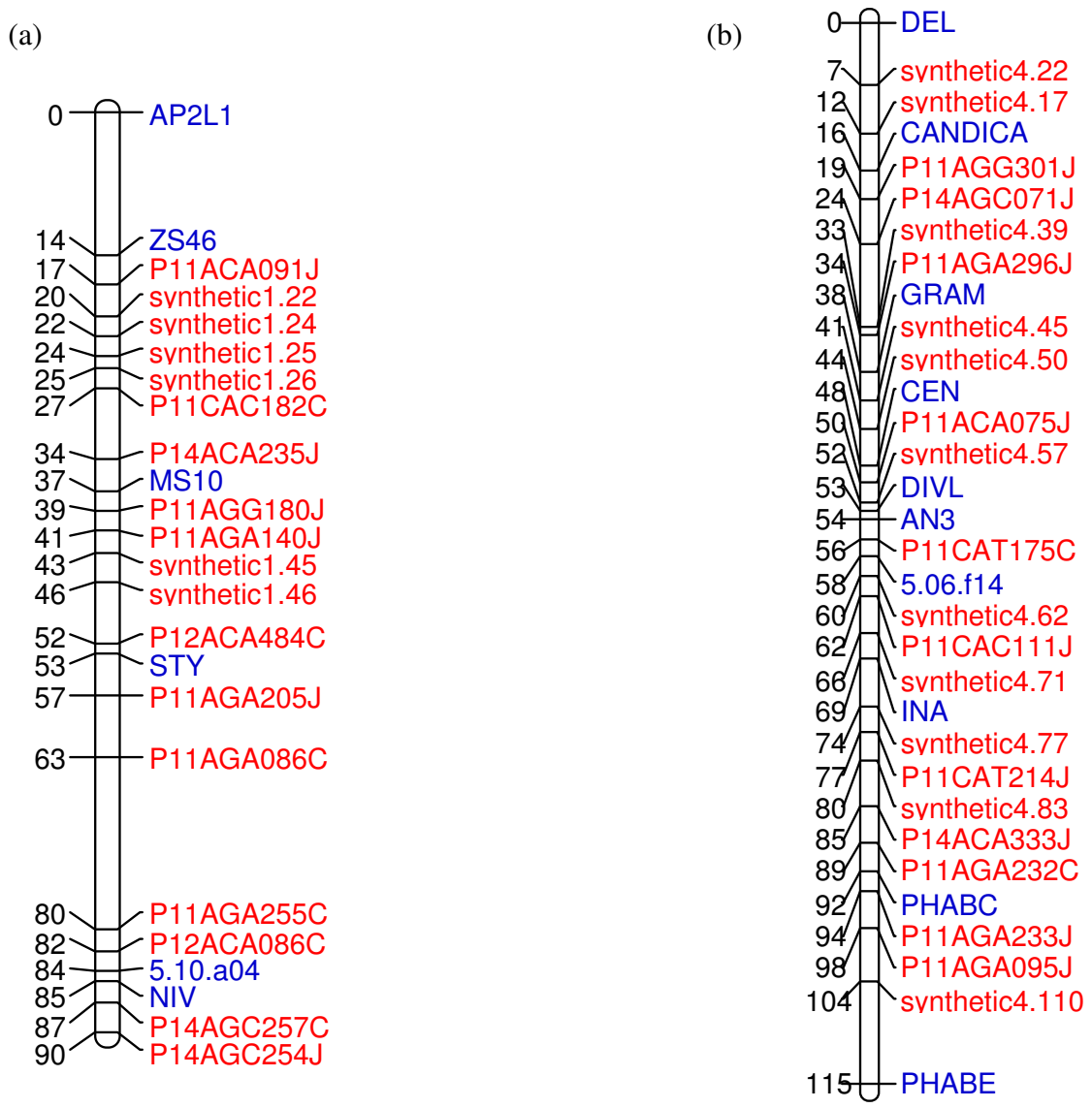


Figure 32. Linkage map position of *Niv*, *Candi* and *Del* (a) The *Niv* gene was mapped in LG1 at position 85 cM, (b) *Candi* in LG4 at 16 cM and *Del* in the same chromosome at position 0 cM. Markers corresponding to protein coding genes are shown in blue and AFLP markers in red. "Synthetic" markers were mapped as co-dominant markers, inferred from the dominant genotypes of AFLPs from both parental chromosomes that were not separated by recombination. Taken from Feng *et al.* (2009)

In an attempt to develop a co-dominant marker for *Niv*, partial first, second and third exon, full first and second intron from *A. majus* and *A. charidemi* were sequenced, but no polymorphism was found (data not shown; primers were designed based on *Niv* gene data, accession number X03710). However, because *Niv* is located in LG1, in which no

flower colour QTL was found, *Niv* seems unlikely to make a major contribution for flower colour variation between *A. majus* and *A. charidemi*. However, I have test the correlation between flower colour and *AmNiv* allele in V1640 F_2 population using the existing dominant marker data from Feng *et al*. (2009); no correlation was found between *AmNiv* allele and lobe (R^2= 0.009, P-value = 0.3513) or tube (R2= 0.009, P-value = 0.3621) variation.

ii. *Chalcone isomerase* (*CHI*)

The second gene in the ABP encodes chalcone isomerase (CHI). CHI catalyzes isomerization of the yellow tetrahydroxychalcone to colourless naringenin. However, even without CHI activity, chalcone can spontaneously isomerize to form naringenin at slower rate (Holton and Cornish, 1995). CHI is now known to exist as two isozymes in some species (Shimada *et al*., 2003; Ralston *et al*., 2005). CHI-encoding cDNA (*CHI*) was first cloned from *A. majus* by Martin *et al*. (1991). There is no report of a type 2 *CHI*-like gene in *A. majus* so far, and no additional *CHI*-like genes could be detected in an EST collection for *A. majus* or the *A. majus* genome (data not shown). *CHI* was reported to be located in LG4 of a map made in an *A. majus* x *A. molle* F_2 hybrid population by Schwarz-Sommer *et al*. (2003). Therefore *CHI* was not considered as a candidate for a major flower colour QTL. However, the expression of *CHI* was compared between pale and dark individuals in M239 family, which will be discussed later in next chapter.

iii. *Incolorata* (*Inc*)

The third gene in the ABP, *Incolorata* (*Inc*) encodes flavanone-3-hydroxylase (F3H). F3H converts naringenin to dihydrokaempferl (DHK; Holton and Cornish, 1995). Mutation in *inc* gene give ivory coloured flowers with yellow faces (Figure 31b), because they are able to produce the substrate naringenin and yellow aurone pigments derived from it. In the previous chapter, *Inc* was mapped to LG7 at 8 cM and had shown

a significant correlation to flower colour variation in the *A. majus* x *A. charidemi* F_2 family V1640 family for both lobe and tube colour, making it a candidate for the major-effect QTL in LG7. A further reason to suspect involvement of *Inc* is that loss-of-function mutations are not fully recessive to wild-type (Stubbe, 1966). This suggests that anthocyanin biosynthesis is sensitive to the expression level of *Inc* and that half the wild-type level of expression in heterozygotes is insufficient for normal levels of anthocyanin pigment. Therefore relatively minor variation in *Inc* activity has the potential to affect flower colour.

Regression analysis suggested that *Inc* can explain about 14% of corolla lobe colour variation and 7% of corolla tube variation in the F_2 family V1640. The NIL family M244 was genotyped at *Inc*, and the correlation between *Inc* genotype and lobe colour was found to have increased (Figure 33a), presumably because fewer QTL segregated in the NIL. Here *Inc* genotypes could explain about 31% of lobe colour variation in M244. However, this does not rule out linkage of the major effect QTL and *Inc*, a view that was supported by segregation of flower colour in family M239, which is homozygous at *Inc* (Figure 25). However, it did support the presence of a major-effect flower colour QTL in LG7 and variation found in M244 family suggested that it was heterozygous in at least one of the NILs families.

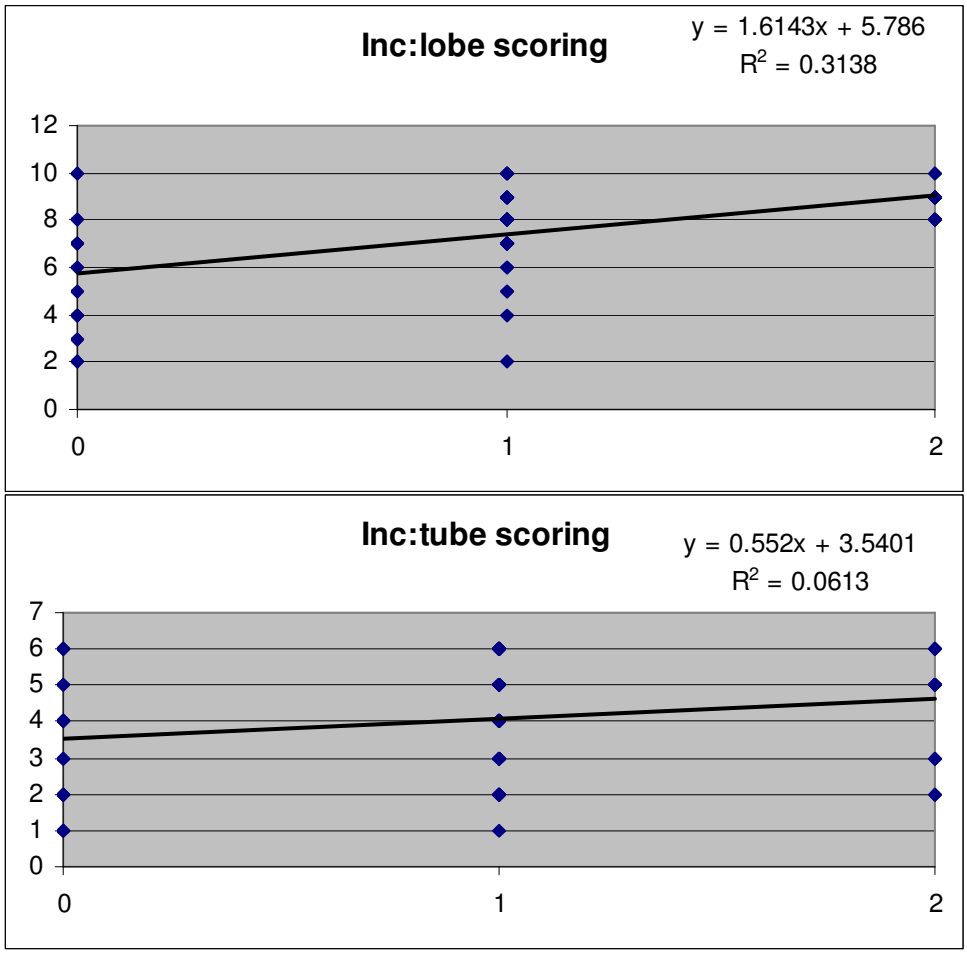


Figure 33. Simple linear regression of colour variation and *Inc* genotypes in NIL family M244 (n = 98) The x-axis shows the dose of the *A. majus Inc* allele, so *A. charidemi* homozygous = 0, heterozygous = 1, *A. majus* (JI7) homozygous = 2, with lobe colour (a) or tube colour (b) each on a scale of 0-10 plotted against allele dose. For lobes, R^2 = 0.3138 and the P-value for colour being correlated with genotype (from one-way *ANOVA*) 1.38 x 10^{-8}. For tube colour, R^2=0.0613 and P=0.02. Since the mean of heterozygotes is approximately half way between the homozygotes, simple linear regression (i.e., assuming no dominance of the *A. majus* allele) is justified.

iv. *Eosina* (*Eos*)

Eosinea (*Eos*) encodes flavonoid-3'-hydroxylase (F3'H). F3'H hydroxylates DHK at the 3'position of the B-ring to produce dihydroquercetin (DHQ; Holton and Cornish, 1995), the substrate for the cyanidin branch of the ABP. *eos* mutants therefore produce only pelargonidins and have a more pinkish colour (Figure 33d). F3'H, along with F3'5'H, has been was reported to control flower hue variation in the Texas wildflower *Phlox drummondii* (Hopkins and Rausher, 2011) but not to affect colour intensity. Although this made it an unlikely candidate for a colour-intensity QTL, the scoring method used for *Antirrhinum* flowers might be affected by hue variation. I therefore decided to test involvement of the *Eos* gene further in NILs families. However, sequencing the third intron of *Eos* locus detected no polymorphism between the *A. majus* and *A. charidemi* alleles present in V1604 (data not shown; primers were designed based on *Eos* cDNA data, accession number DQ272592). and therefore attempts to map *Eos* were not pursued.

v. *Pallida* (*Pal*)

The fifth gene, *Pallida* (*Pal*) encodes dihydroflovonol redutase (DFR). DFR is required to convert colourless dihydroflavonols (DHK and DHQ) to leucoanthocyanidins (leucopelargonidin and leucocyanidin; Holton and Cornish, 1995). Loss of *pal* activity gives an ivory flower with yellow face (Figure 33c), similar to the effects of *chi* and *inc* mutations. *Pal* has been mapped to the same linkage group as *Ros* by Whibley *et al*., (2006) and at 27 cM in LG3 by Schwarz-Sommer *et al.*, (2003), making it a candidate for the major-effect QTL in LG3. Though some species such as Japanese morning glory (*Ipomoea nil)*, Petunia, *Gerbera hybrida* and *Medicago truncatula* were reported to have more than one copy of DFR, normally only one copy is involved in anthocyanin synthesis in flowers (Inagaki *et al*., 1999; Holton and Cornish, 1995; Helariutt *et al*,. 1993; Xie *et al*., 2004). DFR was reported as a single gene in barley (*Hordeum vulgare*), Arabidopsis, tomato (*Lycopersicon esculentum*), grape (*Vitis vinifera*), and rice

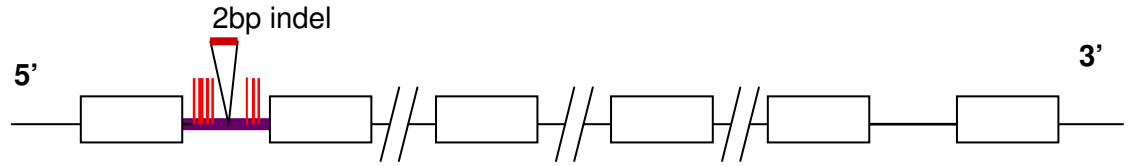
(*Oryza sativa*; Kristiansen and Rohde, 1991; Winkel-Shirley *et al*., 1992; Bongue-Bartelsman *et al*., 1994; Sparvoli *et al*., 1994; Holton and Cornish, 1995; Chen *et al*., 1998; Reviewed by Xie *et al*., 2004). Similarly, in *A. majus*, only one single copy of *Pal* has been found (Martin *et al*., 1991). *DFR* genes have been reported to underlie natural variation in flower colour in Petunia, morning Glory and Gerbera (van der Krol *et al*, 1990; Inagaki *et al*., 1999; Helariutt *et al*,. 1993). Therefore the involvement of *Pal* in *Antirrhinum* natural variation was further investigated in NILs families.

The *Antirrhinum Pal* (*AmPal*) gene has five introns and six exons (Figure 34a). For allelic differentiation, the *Pal* alleles from two different lines of *A. majus* were sequenced (*JI7Pal* from line JI.7 and *JI98Pal* from line JI:98) along with alleles from two accessions of *A. molle* (*AmoPal1* and *est57Pal*) and *A. charidemi* (*AcPal)* (primers were designed based on *Pal* gene data, accession number X15536). Three regions of *Pal* were compared, the first intron, the final intron and part of the final exon (Figure 35, 36). No difference was found between the alleles from the two *A. molle* accessions and therefore the *A. molle* genotype is referred to as *AmoPal*.

In the first intron, the *JI98Pal* allele differed from all the other alleles at two polymorphic sites (the first involving two adjacent SNPS, the second three adjacent SNPs; Fig. 34b). *AmoPal* differed from the other alleles at an indel (insertion or deletion) of two nucleotides and a further three nucleotide substitutions (Figure 34b).

In the final intron, the two *A. majus* sequences were identical, as were the two *A. molle* sequences. No difference was found between the sequences from *A. majus* and *A. charidemi* (*AmPal* and *AcPal*), but both differed from the *A. molle* allele, *AmoPal*, by one indel and three substitutions (Figure 35).

(a)



(b)

```
JI7Pal          GTCATCAGCAGCCACGATGGTCAATAATCGACGAATAGTTAGTTAGACTGCAACACTGAC 60
JI98Pal         GTCATCAGCAGCCACGATGGTCAATTTTCTCGGAATAGTTAGTTAGACTGCAACACTGAC 60
AcPal           GTCATCAGCAGCCACGATGGTCAATAATCGACGAATAGTTAGTTAGACTGCAACACTGAC 60
AmoPal          GTCATCAGCAGCCACGATGGTCAATAATCGACGAATAGTTAGTTAGACTGCAACACTGAC 60

JI7PAL          AAACCTACAACTATCTTTTCTCAATAGACTCAAGTCATGACTCTCTTTTCTCAATAGACT 120
JI98PAL         AAACCTACAACTATCTTTTCTCAATAGACTCAAGTCATGACTCTCTTTTCTCAATAGACT 120
AcPal           AAACCTACAACTATCTTTTCTCAATAGACTCAAGTCATGACTCTCTTTTCTCAATAGACT 120
AmoPal          AAACCTACAACTATCTTTTCTCAATAGACTCAAGTCATGACTCTCTTTTCTCAATAGACT 120

JI7PAL          CAAGTCATGACTCCCTTTTCGACGTAGAGAATTAGCAACAATTCCAAAT--AAAAATGTG 178
JI98PAL         CAAGTCATGACTCCCTTTTCGACGTAGAGAATTAGCAACAATTCCAAAT--AAAAATGTG 178
AcPal           CAAGTCATGACTCCCTTTTCGACGTAGAGAATTAGCAACAATTCCAAAT--AAAAATGTG 178
AmoPal          CAAGTCATGACTCCCTTTTCGACGTAGAGAATTAGCAACAATTCCAAATTAAAAAATGTG 180

JI7PAL          TAGTCATACCAGGATCACGAACAGTTGCACGAACGGTGTAGCCACGTTCGAGAAGCCTCA 238
JI98PAL         TAGTCATACCAGGATCACGAACAGTTGCACGAACGGTGTAGCCACGTTCGAGAAGCCTCA 238
AcPal           TAGTCATACCAGGATCACGAACAGTTGCACGAACGGTGTAGCCACGTTCGAGAAGCCTCA 238
AmoPal          TAGTCATACCAGGATCACGAACAGTTGCACGAACGGTGTAGCCACGTTCAAGAAGCCTCA 240

JI7PAL          TGACAAGCCATGAGCCTATGAAGCCGGCAGCTCCTGTGACACATACGGTGGAA 291
JI98PAL         TGACAAGCCATGAGCCTATGAAGCCGGCAGCTCCTGTGACACATACGGTGGAA 291
AcPal           TGACAAGCCATGAGCCTATGAAGCCGGCAGCTCCTGTGACACATACGGTGGAA 291
AmoPal          TAACAAGCCATGAGCCGATGAAGCCGGCAGCTCCTGTGACACATACGGTGGAA 293
```

Figure 34.Variation in the first intron of *Pal* within *Antirrhinum* (a) A schematic diagram showing the structure of *Pal* locus, with five introns and six exons (not to scale). The first intron that was sequenced is shown as a thick purple line. Boxes indicate exon, lines indicates non-coding DNA. Polymorphisms in the first intron are shown in red, with SNPs as line and an indel as a triangle. (b) Sequences of the first intron from different *Antirrhinum* accessions (*JI7Pal*, *A. majus* line JI.7; *JI98Pal*, *A. majus* line JI.98; *AcPal*, *A. charidemi*; *AmoPal*, *A. molle*). Invariant nucleotides are shaded.

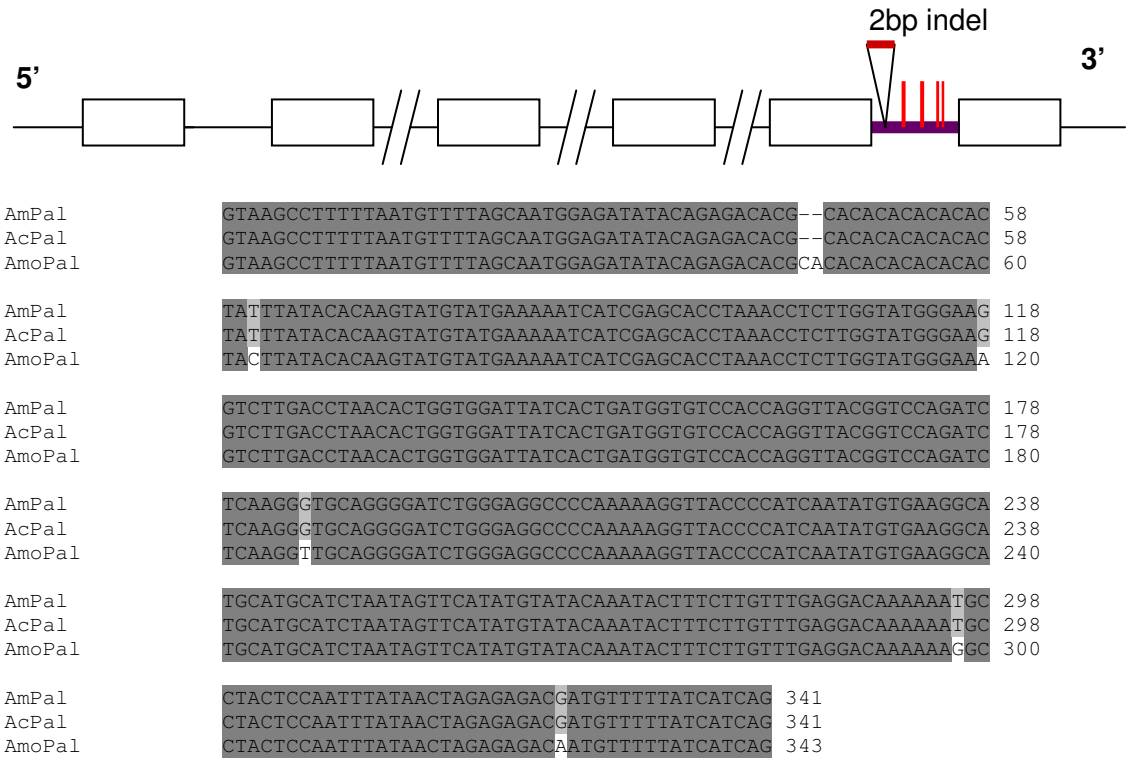


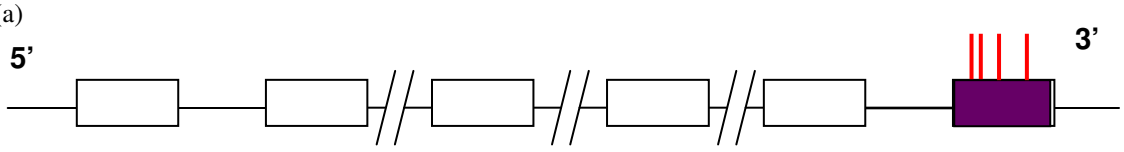
Figure 35. Variation in the final intron of *Pal* within *Antirrhinum* The figure is labelled as for Fig. 34.

Sequencing the last exon of *Pal* identified four SNPs (Figure 36a, b). All four SNPs caused amino acid substitutions to the DFR peptide. Compared to the published *Pal* cDNA sequence from line JI.552 (Martin *et al*., 1991), two substitutions, at cDNA positions 986 and 1291, were found in all other alleles (Figure 36a). They cause amino acid substitution at positions 324 and 425 of the full length protein, from Glycine (G) in *AmPal* to Threonine (T), and from Histidine (H) to Glutamine (Q), respectively (Figure 36c). A SNPs at (cDNA) position 1022 was found in *JI98Pal* and the *A. molle* allele *est57Pal* relative to all other alleles, causing an amino acid substitutions at position 336 from Methionine (M) to Leucine (L) in *A. molle* est57 (Fig. 36c). An additional SNP at position 1095 distinguished the *JI7Pal* allele from all others, causing an amino acid substitution from Tyrosine (Y) to Serine (S) at position 360 in *JI7Pal*. Surprisingly, the data suggest that variation between lines of the same species can be greater than that between species. This suggests that these mutations do not affect DFR activity, because they are present in *A. majus*, which has uniformly dark flowers. For both *A. majus*

inbred lines and both *A. molle* lines collected from the field, more than one allele was present within the species, each encoding a different amino acid sequence. However, none of these changes seem likely to cause a complete loss of DFR function, because all the lines can produce anthocyanin in flowers (*A. majus* has dark flowers, *A. charidemi* pale pink flowers and *A. molle* produces anthocyanin over veins).

NIL family M239 family was genotyped for *Pal* by sequencing, to test whether segregation of *Pal* contributed to the flower colour variation in the family. All individual of M239 were homozygous for the *AcPal* allele, and therefore *Pal* could not contribute to variation in this NIL. This does not rule out the possibility that *Pal* represents a flower colour QTL that was not heterozygous in the NIL.

(a)



(b)

```
cDNAPal          TGACAAGGACATACCGGTGGTGAGTTTCTCCTCCAAGAAAATGATCGGAATGGGTTTTAT 960
JI7Pal                                                    CGGAATGGGTTTTAT 15
JI98Pal                                                   CGGAATGGGTTTTAT 15
AcPal                                                     CGGAATGGGTTTTAT 15
AmoPal                                                    CGGAATGGGTTTTAT 15
est57Pal                                                  CGGAATGGGTTTTAT 15

cDNAPal          ATTCAAGTACACATTAGAGGACATGGTCAGGGGTGCGATTGACACTTGCCGCGAGAAAGG 1020
JI7Pal           ATTCAAGTACACATTAGAGGACATGTTCAGGGGTGCGATTGACACTTGCCGCGAGAAAGG 75
JI98Pal          ATTCAAGTACACATTAGAGGACATGTTCAGGGGTGCGATTGACACTTGCCGCGAGAAAGG 75
AcPal            ATTCAAGTACACATTAGAGGACATGTTCAGGGGTGCGATTGACACTTGCCGCGAGAAAGG 75
AmoPal1          ATTCAAGTACACATTAGAGGACATGTTCAGGGGTGCGATTGACACTTGCCGCGAGAAAGG 75
est57Pal         ATTCAAGTACACATTAGAGGACATGTTCAGGGGTGCGATTGACACTTGCCGCGAGAAAGG 75

cDNAPal          AATGCTCCCGTATTCTACTAAGAATAACAAGGGAGACGAGAAGGAACCCATTTTGAATTC 1080
JI7Pal           AATGCTCCCGTATTCTACTAAGAATAACAAGGGAGACGAGAAGGAACCCATTTTGAATTC 135
JI98Pal          ATTGCTCCCGTATTCTACTAAGAATAACAAGGGAGACGAGAAGGAACCCATTTTGAATTC 135
AcPal            AATGCTCCCGTATTCTACTAAGAATAACAAGGGAGACGAGAAGGAACCCATTTTGAATTC 135
AmoPal1          AATGCTCCCGTATTCTACTAAGAATAACAAGGGAGACGAGAAGGAACCCATTTTGAATTC 135
est57Pal         ATTGCTCCCGTATTCTACTAAGAATAACAAGGGAGACGAGAAGGAACCCATTTTGAATTC 135
```

Figure 36. Allelic variation in the final exon of *Pal* (a) The region of the final exon sequenced, with SNPs shown by red bars. Introns and exons are shown as in the previous two figures. (b) Nucleotide sequence comparisons. The published cDNA sequence from JI.552 (sequence accession GI:16026) is shown underlined in blue. Sequence of the alleles from *A. majus* lines *JI.7Pal* and *JI.98Pal*, from two accessions of *A. molle*, *AmoPal1* and *est57Pal*, and from *A. charidemi*, *AcPal*, are shown. Invariant nucleotides are shaded (to be continued).

(b; continued)

```
cDNAPal          CCTAGAAAACAATTACAATATTCAAGACAAAGAACTGTTTCCAATTTCGGAGGAAAAACA 1140
JI7Pal           CCTAGAAAACAATTCCAATATTCAAGACAAAGAACTGTTTCCAATTTCGGAGGAAAAACA 195
JI98Pal          CCTAGAAAACAATTACAATATTCAAGACAAAGAACTGTTTCCAATTTCGGAGGAAAAACA 195
AcPal            CCTAGAAAACAATTACAATATTCAAGACAAAGAACTGTTTCCAATTTCGGAGGAAAAACA 195
AmoPal1          CCTAGAAAACAATTACAATATTCAAGACAAAGAACTGTTTCCAATTTCGGAGGAAAAACA 195
est57Pal         CCTAGAAAACAATTACAATATTCAAGACAAAGAACTGTTTCCAATTTCGGAGGAAAAACA 195

cDNAPal          CATCAATGGACAAGAGAATGCCCTGCTTTCAAATACTCAAGACAAAGAACTGCTTCCAAC 1200
JI7Pal           CATCAATGGACAAGAGAATGCCCTGCTTTCAAATACTCAAGACAAAGAACTGCTTCCAAC 255
JI98Pal          CATCAATGGACAAGAGAATGCCCTGCTTTCAAATACTCAAGACAAAGAACTGCTTCCAAC 255
AcPal            CATCAATGGACAAGAGAATGCCCTGCTTTCAAATACTCAAGACAAAGAACTGCTTCCAAC 255
AmoPal1          CATCAATGGACAAGAGAATGCCCTGCTTTCAAATACTCAAGACAAAGAACTGCTTCCAAC 255
est57Pal         CATCAATGGACAAGAGAATGCCCTGCTTTCAAATACTCAAGACAAAGAACTGCTTCCAAC 255

cDNAPal          TTCAGAAGAAAAACGTGTTAATGGACTAGAGAGCGCCCTGCTTTCAAAGATTCAAGACAA 1260
JI7Pal           TTCAGAAGAAAAACGTGTTAATGGACTAGAGAGCGCCCTGCTTTCAAAGATTCAAGACAA 315
JI98Pal          TTCAGAAGAAAAACGTGTTAATGGACTAGAGAGCGCCCTGCTTTCAAAGATTCAAGACAA 315
AmoPal1          TTCAGAAGAAAAACGTGTTAATGGACTAGAGAGCGCCCTGCTTTCAAAGATTCAAGACAA 315
AcPal            TTCAGAAGAAAAACGTGTTAATGGACTAGAGAGCGCCCTGCTTTCAAAGATTCAAGACAA 315
est57Pal         TTCAGAAGAAAAACGTGTTAATGGACTAGAGAGCGCCCTGCTTTCAAAGATTCAAGACAA 315

cDNAPal          AGAAGTGCTTCCAACTTCAGGGGTTAAGCATGCCAAGGGACAAGAAAATGCGCTGCTTCC 1320
JI7Pal           AGAAGTGCTTCCAACTTCAGGGGTTAAGCAA                              346
JI98Pal          AGAAGTGCTTCCAACTTCAGGGGTTAAGCAA                              346
AcPal            AGAAGTGCTTCCAACTTCAGGGGTTAAGCAA                              346
AmoPal1          AGAAGTGCTTCCAACTTCAGGGGTTAAGCAA                              346
est57Pal         AGAAGTGCTTCCAACTTCAGGGGTTAAGCAA                              346
```

(c)

```
Pal              VVSFSSKKMIGMGFIFKYTLEDMVRGAIDTCREKGMLPYSTKNNKGDEKEPILNSLENNY 360
JI7Pal                     GMGFIFKYTLEDMFRGAIDTCREKGMLPYSTKNNKGDEKEPILNSLENNS  50
JI98Pal                    GMGFIFKYTLEDMFRGAIDTCREKGLLPYSTKNNKGDEKEPILNSLENNY  50
AcPal                      GMGFIFKYTLEDMFRGAIDTCREKGMLPYSTKNNKGDEKEPILNSLENNY  50
AmoPal1                    GMGFIFKYTLEDMFRGAIDTCREKGMLPYSTKNNKGDEKEPILNSLENNY  50
est57Pal                   GMGFIFKYTLEDMFRGAIDTCREKGLLPYSTKNNKGDEKEPILNSLENNY  50

Amino acid seq   NIQDKELFPISEEKHINGQENALLSNTQDKELLPTSEEKRVNGLESALLSKIQDKEVLPT 420
JI7Pal           NIQDKELFPISEEKHINGQENALLSNTQDKELLPTSEEKRVNGLESALLSKIQDKEVLPT 110
JI98Pal          NIQDKELFPISEEKHINGQENALLSNTQDKELLPTSEEKRVNGLESALLSKIQDKEVLPT 110
AcPal            NIQDKELFPISEEKHINGQENALLSNTQDKELLPTSEEKRVNGLESALLSKIQDKEVLPT 110
AmoPal1          NIQDKELFPISEEKHINGQENALLSNTQDKELLPTSEEKRVNGLESALLSKIQDKEVLPT 110
est57Pal         NIQDKELFPISEEKHINGQENALLSNTQDKELLPTSEEKRVNGLESALLSKIQDKEVLPT 110

Amino acid seq   SGVKHAKGQENALLPDIANDHTDGRI  446
JI7Pal           SGVKQ                      115
JI98Pal          SGVKQ                      115
AcPal            SGVKQ                      115
AmoPal1          SGVKQ                      115
est57Pal         SGVKQ                      115
```

Figure 36 (continued). Allelic variation in the final exon of *Pal* (a) The region of the final exon sequenced, with SNPs shown by red bars. Introns and exons are shown as in the previous two figures. (b) Nucleotide sequence comparisons. The published cDNA sequence from JI.552 (sequence accession GI:16026) is shown underlined in blue. Sequence of the alleles from *A. majus* lines *JI.7Pal* and *JI.98Pal*, from two accessions of *A. molle*, *AmoPal1* and *est57Pal*, and from *A. charidemi*, *AcPal*, are shown. Invariant nucleotides are shaded. (c) Amino acid translations of the alleles shown in (b). The JI.552 sequence is shown in green, invariant amino acids are shaded.

vi. *Candica (Candi)*

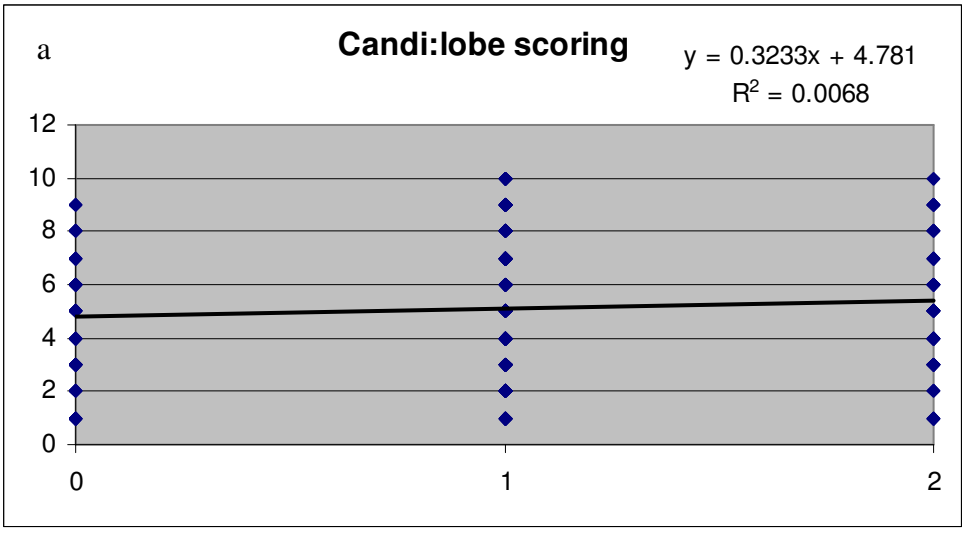
The *Candica* (*Candi*) locus, which encodes anthocyanidin synthase (ANS), was cloned by Martin *et al*., (1991). This enzyme is thought to convert leucoanthocyanins to anthocyanidins (both pelargonidin and cyanidin) by oxidation and dehydration (Bartlett, 1989; Heller and Forkmann, 1988). Flowers of *candi* mutants resemble *Pal* mutants, with ivory flowers and yellow faces (Figure 31e).

Candi was mapped to 16 cM in LG4 in family V1640 by Feng *et al*. (2009; Figure 32b), so it was unlikely to correspond to one of the two major effect QTL. However, a minor effect QTL for tube colour had been detected in LG4. Therefore, existing genotype data for *Candi* was used in regression analysis with lobe and tube colour with tube and lobe colour scores (Figure 37a, b). One-way *ANOVA* suggested that tube colour is significantly correlated with *Candi* genotype (P =0.0009). Regression analysis suggested that about 6% of the variation found in V1640 could be explained by *Candi*.

vii. *Anthocyanin Glycosyltransferases gene (UFGTs)*

Glycosyltransferases (GTs) is one class of enzymes that modify anthocyanin, performing the specific glycosylation at hydroxyls using UDP-sugars (Gachon, Langlois-Meurine and Sainrenan, 2005). The sugar conjugation results in increasing stability and water solubility and the glycosylated form of anthocyanins, which can also be transferred more easily from cytoplasm to the vacuole (Yonekura-Sakakibara *et al*., 2009). The most common glycosylation involves modification of the 3-hydroxyl by UDP-glucose:anthocyanidin 3-*O*-glucosyltransferase (3GT), except in rose (Ogata *et al*., 2005).

A cDNA for a putative UDP glucose:flavonoid 3-*O*-glucosyltransferase gene (*UFGT*) called pJAM338 was isolated from snapdragon by Martin *et al*. (1991). It encoded a



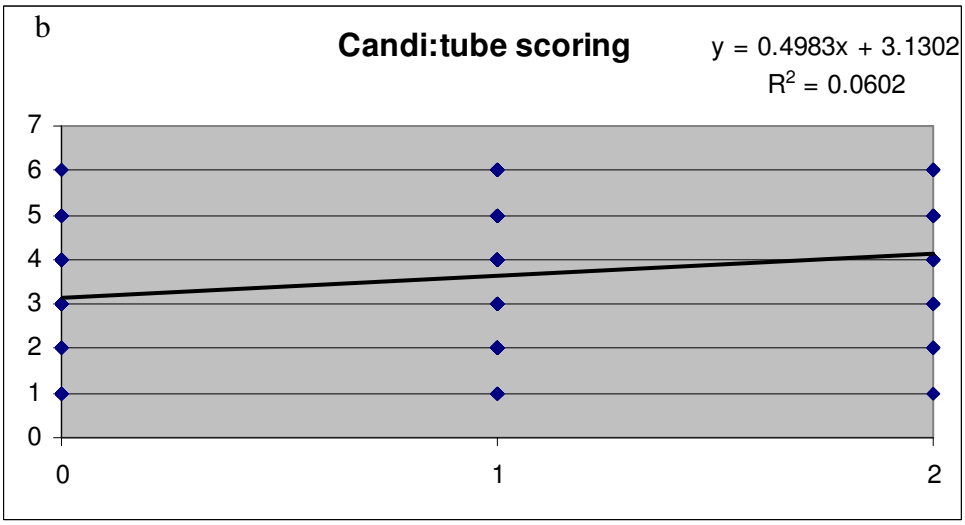


Figure 37. Simple linear regression analysis of tube and lobe colour with *Candi* genotypes in the V1640 F_2 population (n = 140). Dosage of the *A. majus Candi* allele is shown on the y-axis and colour score on the y-axis. Lobe colour (a) is not significantly correlated with genotype, R^2=0.0068, P=0.27. Tube colour shows a significant correlation to *Candi* genotype, R^2 = 0.0602, P=0.0009.

protein with about 58% amino acid similarity to maize UFGT. Schwinn and colleagues (1993) tested its activity in *Lisianthus* and confirmed that it has 3GT activity. The sequence was mapped to 11 cM in LG2 in an *A. majus* x *A. molle* F_2 population (Schwarz-Sommer *et al*., 2003), and was therefore not an obvious candidate for any of the QTL detected in the *A. majus* x *A. charidemi* population. Unfortunately, time did not permit further investigation of this gene. Thus, the possibility that *UFGT* is one of the small QTLs for lobe (62 cM LG2) or tube (88 cM LG2) variation, still remains.

viii. *Anthocyanidin rhamnosyltransferase gene (3RT)*

Antirrhinum species normally produce anthocyanin of two types, cyanindins and pelargonidins, but are unable to produce delphinidins due to the lack of dihydroflavonol-3,5-hydroxylase. The major anthocyanins in *Antirrhinum* are thought to be cyanidin-3-rutinoside and pelargonidin-3- rutinoside (Harborne, 1963).

To produce anthocyanidin-3-rutinoside from anthocyanidin-3-glucosides, the enzyme UDP rhamnose:anthocyanidin-3-glucoside rhamnosyltransferase (3RT) is required (Holton and Cornish, 1995). The *3RT* gene (*Rt* locus) was cloned from Petunia using a combination of differential screening and genetic mapping (Brugliera *et al*., 1994; Kroon *et al*., 1994) and found to show a low level of sequence homology to *UFGT* and other glycosyl transferase (Holton and Cornish, 1995). However there is no evidence of a more similar sequence than *UFGT* in snapdragon.

4.2.2 Other structural genes

i. *Anthocyanin transporter gene/s*

Epidermal cells concentrate anthocyanins in the vacuole, so the efficiency of transporting anthocyanin into vacuole could affects flower colour intensity. Glutathione transferase (GST) is believed to affect this process, possibly by forming an anthocyanin-GST complex that protects anthocyanin. However, the mechanism of transport is still unclear (Zhao *et al*, 2010). In *Antirrhinum*, no glutathione transferase has been reported

to be associated with anthocyanin transport. Recently, an anthocyanin permease gene (*PAT*) was identified in tomato and shown to be up-regulated in transgenic fruits that over-expressed a regulator of the ABP (Mathews *et al.*, 2003; Butelli *et al.* 2008), its sequence was homologous to *TT12* in Arabidopsis (Marinova *et al.,* 2007) and *MATE1* in *Medicago truncatula* (Zhao and Dixon, 2009), suggesting it was the vacuole-localized multidrug and toxic compound extrusion (MATE) transporter. There was therefore the possibility that a related gene might have a function in anthocyanin accumulation in *Antirrhinum* corolla cells.

Blasting ESTs sequences from *A. majus* (www.Antirrhinum.net) with the anthocyanin permease (PAT) gene from tomato and *Vitis* identified four ESTs with some similarity (AJ796511, AJ795263, AJ806743 and AJ794241, Fig. 38). These were named *PAT1*, *PAT2*, *PAT3* and *PAT4*, respectively. Because the *A. majus* genome sequence was not available at that time, they were aligned with the grape *PAT* genomic sequence (accession number LOC100268149) to estimate the likely positions of intron and exon in the gene (Figure 38). Primers to amplify regions of the *Antirrhinum PAT* genes were designed from these alignments. Unfortunately, *PAT3* could not be amplified from *A. majus* or *A. charidemi*. *PAT1* and *PAT4* were amplified and sequenced, but no polymorphisms were found between the alleles from *A. majus* and *A. charidemi* (data not shown). One SNP was found between the alleles of *PAT2* fragments (Figure 39). However, when expression of *PAT2* and *PAT4* was compared by RT-PCR in pale and dark individuals of NIL family M239, neither showed detectable expression, as discussed later in the chapter. This suggests that *PAT2* and *PAT4* are not involved in anthocyanin accumulation in flowers.

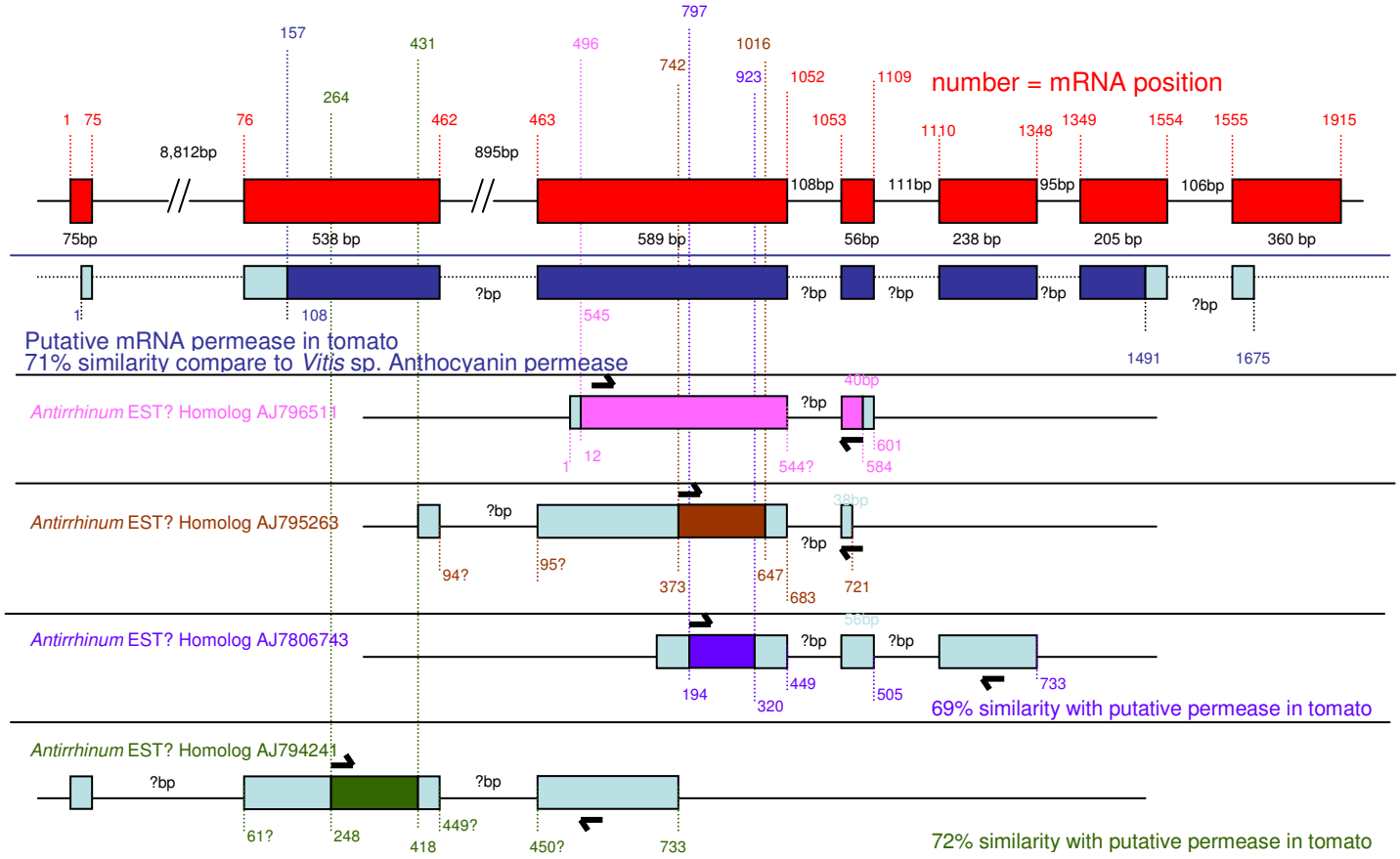
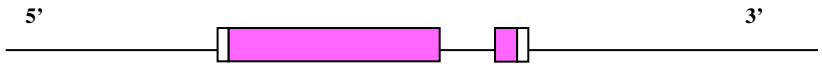


Figure 38. Putative *PAT* loci in *Antirrhinum* EST Schematic diagram show the alignment of the grapevine permease genomic sequence (red; accession LOC100268149) with the cDNA sequence of a tomato anthocyanin permease gene (*PAT*; blue; accession number AY348872) and four related *Antirrhinum* ESTs (from www.Antirrhinum.net; AJ796511 (pink), AJ795263 (brown), AJ806743(blue) and AJ794241(green). Positions of the primers designed to amplify coding and non-coding regions of the *Antirrhinum* sequences are shown with black arrows. Primers sequences are given in Table 2. Coloured boxes show the regions that have high similarity to *PAT* coding sequences from tomato and grapevine.




```
AmPAT2          GTGAGGAGACACTAGCACTGAGAATTTCGCCGCTTTTGCATTGCCAGCTCCAAGTTCATT 60
AcPAT2          GTGAGGAGACACTAGCACTGAGAATTTCGCCGCTTTTGCATTGCCAGCTCCAAGTTCATT 60

AmPAT2          CGCCACTCTCACACTGAAAATCAAATACACAAATGTTAAATATGTAGTAAACAGGTATTA 120
AcPAT2          CGCCACTCTCACACTGAAAATCAAATACACAAATGTTAAATATGTAGTAAACAGGTATTA 120

AmPAT2          AAATCAACTATTAGGCAGTTGGAACTTTACAGATTCTGAGTTAGTATTGAAAAAAAAAAG 180
AcPAT2          AAATCAACTATTAGGCAGTTGGAACTTTACAGATTCTGAGTTAGTATTGAAAAAAAAAAG 180

AmPAT2          ATTAGGAATACATAAAGTAAGAAACATCACCTTATGGCAGCATTGAATCCAATTGAGATC 240
AcPAT2          ATTAGGAATACATAAAGTAAGAAACATCCCCTTATGGCAGCATTGAATCCAATTGAGATC 240

AmPAT2          ATGGCATCCCATCCATTAATATTCATGCTGCATCGGAATGTTGCAGTGACAAAGTTAATC 300
AcPAT2          ATGGCATCCCATCCATTAATATTCATGCTGCATCGGAATGTTGCAGTGACAAAGTTAATC 300

AmPAT2          CAAGTATCACGATGTATATTACGATGGTCTAAACGATTTATCTTGCTTTATTTGAATTAA 360
AcPAT2          CAAGTATCACGATGTATATTACGATGGTCTAAACGATTTATCTTGCTTTATTTGAATTAA 360

AmPAT2          ATGAGACAGAAAAAATGGAATTTAAGGAAGTGACATACCAGATAGAAATAGCATCAACAG 420
AcPAT2          ATGAGACAGAAAAAATGGAATTTAAGGAAGTGACATACCAGATAGAAATAGCATCAACAG 420

AmPAT2          GTATGACAGGGTTATCCAAGTGACCTGCTATAACCACTATAACCATCAAGTACCAAAACT 480
AcPAT2          GTATGACAGGGTTATCCAAGTGACCTGCTATAACCACTATAACCATCAAGTACCAAAACT 480

AmPAT2          CCAAGCTGCATTTTGAGTATTAACCATTAATCTCACAAATATGAATCACAAGTAACAAAC 540
AcPAT2          CCAAGCTGCATTTTGAGTATTAACCATTAATCTCACAAATATGAATCACAAGTAACAAAC 540

AmPAT2          GTAATCGAGCTTTACAAAAAAAACTCTAACCATAACATTATAGCAGAGCCGGCGAGAGT 598
AcPAT2          GTAATCGAGCTTTACAAAAAAAACTCTAACCATAACATTATAGCAGAGCCGGCGAGAGT 598
```

Figure 39. Allelic variation at *PAT2* Partial *PAT2* (EST AJ796511) sequence showing one SNP between *A. majus* (*AmPAT2)* and *A .charidemi (AcPAT2*).

4.2.3 Flavonoid-anthocyanin competition; *Flavonoid reductase gene (FLS)*

The dihydroflavonols (i.e., DHK, DHQ) are substrates for several enzyme, DFR, FLS F3'H and F3'5'H. In normal condition, anthocyanins and flavonols are synthesized within the same cells, and the enzymes localized to the same cellular compartment. Currently there is no clear explanation how the flow of dihydroflavenols into the flavonoids and anthocyanin pathways is regulated. Holton and Cornish (1995) had been proposed the idea of competition between flavonoid-metabolizing enzymes, with evidence suggesting that accumulation of either flavonoid or anthocyanidin occurred at the other's expense. Therefore it is possible that differences between the relative flux into anthocyanin and flavonoid production can cause flower colour variation. For this

reason, the gene encoding flavonoid synthase (FLS) was tested for its association to flower colour QTL.

Primers were designed from the *Antirrhinum FLS* sequence (accession number; DQ272591) published by Schwinn *et al*. in 2006. The gene contains four exons and three introns. The amplified fragment was expected to contain the second intron, and to be more than 350 bp long. The size of amplified fragment was about 900 bp (magenta boxes and line in Figure 40b). Its identity was confirmed by Blasting (nucleotide Basic Local Alignment Search Tool) against the NCBI and EMBL nucleotide databases and against the *A. majus* genome sequence, once this became available.

A. majus and *A. charidemi* alleles were found to differ at seventeen SNPs in the second intron of *FLS* (Figure 40b). Due to time restrictions, involvement of this gene in natural variation was limited to a test of expression levels, described later in this chapter.

(a)

(b)

```
AmFLS           AGTTTTGGCCTAAGAATCCACCCTCTTACAGGTAATATTTTTGGTTGACTTGCATTTGTA 60
AcFLS           AGTTTTGGCCTAAGAATCCACCCTCTTACAGGTAATATTTTTGGTTGACTTGCATTTGTA 60

AmFLS           AATAAAGAAAATAATAATTAGTTTGTTTTGCAAATTTGAGATTTATTAAAAAATCTAAAC 120
AcFLS           AATAAAGAAAATAATAATTAGTTTGTTTTGCAAATTTGAGATTTATTAAAAAATCTAAAC 120

AmFLS           TTCATCTCAAGTCTACTTTGACAAATTTGTATTGATAATTTATGTATGTAAGACTCAAAT 180
AcFLS           TTCATCTCAAGTCTACTTTGACAAATTTGTATTGATAATTTATGTATGTAAGACTCAAAT 180

AmFLS           TTCATAATCCGAAAAACTAAATAATGTGAGGATATATAGTTCCTTTTTTTTTTTTTGTGT 240
AcFLS           TTCATAATCCGAAAAACTAAATCATGTGAGGATAGATCGTTCCTTTTTTTTTTTTTGTGT 240

AmFLS           GTGTGTGTTTTTGCCAAAAAGAGTTTGAAACTGAGATTTAGTGAGAATTTATCGGGTTAT 300
AcFLS           GTGTGTGTTTTTGCCAAAAAGAGTTTGAAACTGAGATTTAGTGAGAATTTATCGGGTTAT 300

AmFLS           TCATGTAATGTTGAACGCCACCGACCCGCCTCAATATAGTAAGGAACTAAACGAAGGTTT 360
AcFLS           TCATGTAATGTTGAACGCCACCGACCCGCCTCAATATAGTAAGGAACTAAACGAAGGTTT 360

AmFLS           GTTTTATGTTTTATGCCTCATTCACTCACATGTCATATCATTTTTTAAGAGCATGTTATT 420
AcFLS           GTTTTATGTTTTATGCCTCATTCACTCACATGTCATATCATTTTTTAAGAGCATGTTATT 420

AmFLS           TTCTTATGTTATGTAAACTCGCAATCACTATCTTTAAGGTGCGTATTTAGTCAAACCTTC 480
AcFLS           TTCTTATGTTATGTAAACTCGCAATCACTATCTTTAAGGTGCGTATTTAGTCAAACCTTC 480

AmFLS           AAGCTAGACACATTAGTTTGCAAACAACACCAGACAGATAAACCACACTAGACAAACCTA 540
AcFLS           AAGCTAGACACATTAGTTTGCAAACAACACCAGACAGATAAACCACACTAGACAAACCTA 540

AmFLS           GTGTGACAAACTCAAACTTAAAAAGCATACAGTTGTATATTGCTTTCCGTGGGAAACAAA 600
AcFLS           GTGTGACAAACTCAAACTTAAAAAGCATACAGTTGTATATTGCTTTCCGTGGGAGACACA 600

AmFLS           TTTGAGACAAATAAAAAGAATTTGACCACTTCCAACAGCACATAAGTTATGTTCATATAT 660
AcFLS           TTTGAGACGAATAAAAAGACTTTGACCAGGTCCTACGGCACGTAAGTTATGATCAGGTAT 660

AmFLS           ATATATATATCGATTCCACTTTCCAAAAAATTAAAAAAACTAAAGTTGGATTAAAATGGA 720
AcFLS           ATATATACATCGATCCCGCTTTCCAAAAAATTAAAAAAACTAAAGTTGGATTAAAATGGA 720

AmFLS           AAGAAACAAGACTTCTTTCTTCTTTGTCTTTTCCTTTTCTAATCACTGTTCATTAATGTC 780
AcFLS           AAGAAACAAGACTTCTTTCTTCTTTGTCTTTTCCTTTTCTAATCACTGTTCATTAATGTC 780

AmFLS           ATGATTAGCACACTTATTGATTAAACTTATCATTAATTTAAATATTAATAATGGCAGAGA 840
AcFLS           ATGATTAGCACACTTATTGATTAAACTTATCATTAATTTAAATATTAATAATGGCAGAGA 840

AmFLS           AGCGAACGAGGAGTACACAAAGAAGATTAGGGACCGTGTCCGACAAGGTGACGAA 895
AcFLS           AGCGAACGAGGAGTACACAAAGAAGATTAGGGACCGTGTCCGACAAGGTGACGAA 895
```

Figure 40. Allelic variation in *FLS* in Antirrhinum (a) Diagram showing *FLS* gene structure in *Antirrhinum*. Blasting the sequence against the *A. majus* genome suggested that *FLS* in *Antirrhinum* is in the small gene family. However, the other loci did not have any report to confirm their activity. This FLS gene contains four exons and three introns. Purple lines indicate the regions sequenced in this experiment and SNPs are shown with red line. (b) Nucleotide sequences from the second intron of *FLS* of *A. majus* (*AmFLS*) and *A. charidemi* (*AcFLS*) differ by 17 SNPs. Sequences extended into the second and third exons, however no SNPs were found in coding regions. Invariant nucleotides are shaded.

4.2.4 Regulatory genes

The pigmentation and its patterning in flowers does not depend only on structural genes but also their control by regulatory genes. Studies of flower pigmentation in different species also suggested the possibility that its variation in nature could occur more frequently by regulatory gene mutation than by structural genes mutation (Streisfeld and Rausher, 2009; Quattrocchio *et al*., 2006; Quattrocchio *et al*., 1999; Scwhinn *et al*., 2006). Regulatory mutations have been suggested to have fewer pleiotropic effects (Coberly and Rausher, 2008) and they have the potential to affect both the level and pattern of pigmentation. In snapdragon, mutations in a number of regulatory genes controlling floral pigmentation have been described, including *delila* (*del*), *Eluta* (*El*), *rosea* (*ros*), *Venosa* (*Ve*) and *mutabilis* (*mut*) (Wheldale, 1907; Baur, 1910a, 1910b; Kuckuck and Schick, 1930; Stubbe, 1966; all reviewed by Schwinn *et al*., 2006).

The regulatory genes identified at the sequence level so far fall into three functional groups; bHLH, MYB and WD40. These three classes of proteins are thought to form a transcription factor complex controlling late biosynthetic genes (LBGs) in the ABP. The recent model suggests that two bHLHs form a dimer that recruits RNA polymerase II, whereas the MYB contacts a specific recognition site in the promoter DNA. The WD protein is required as a bridge to complex the bHLH dimer and MYB proteins together (Figure 41; Pathirana, 2009).

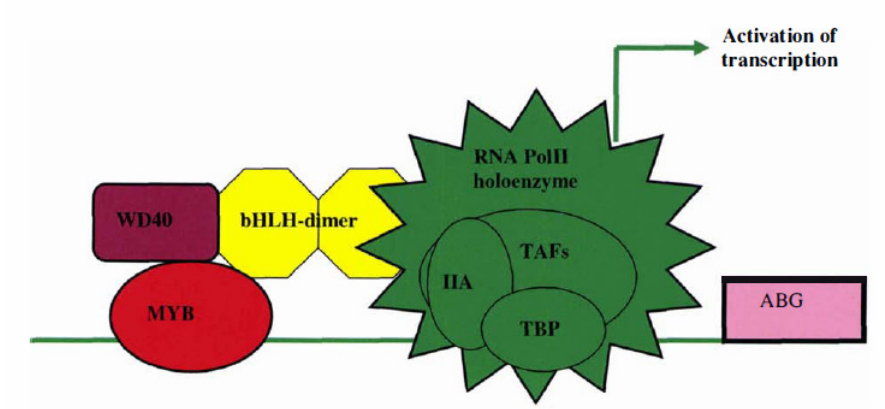
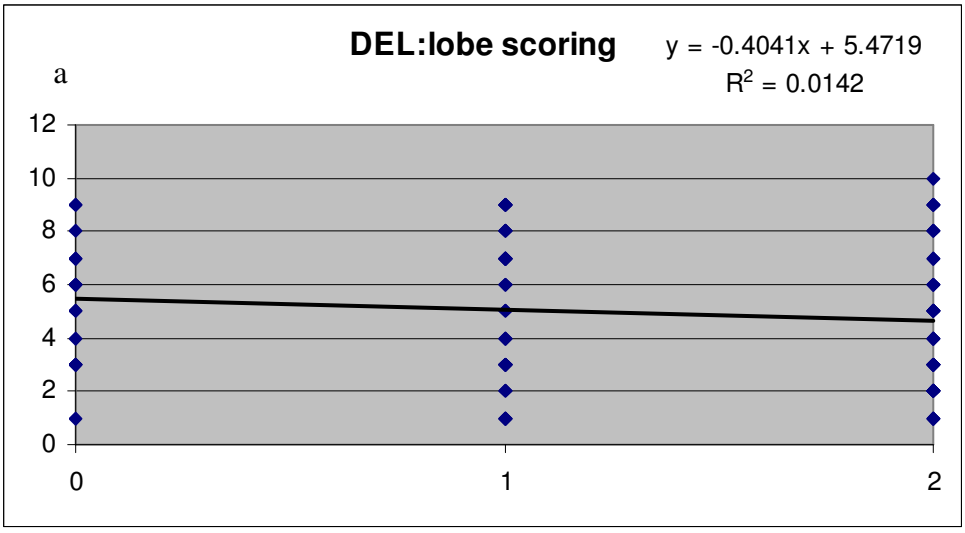


Figure 41. A current model for the initiation of transcription from anthocyanin biosynthesis genes
A complex containing MYB, bHLH and WD40 proteins binds to the promoter regions of the anthocyanin biosynthesis genes and interacts with the basal transcription machinery to activate transcription. Not all the components of the basal transcription machinery are shown, for clarity (from Pathirana, 2007).

i. bHLH regulatory genes

a) *Delila* (*Del*)

Delila (*Del*) was the first regulatory gene from *Antirrhinum* that was well characterized at the molecular level. The recessive *del* mutant prevents anthocyanin biosynthesis in the corolla tube but allows fully pigmented lobes (Goodrich *et al*., 1992; Figure 33f). It was cloned by transposon tagging and identified as encoding a bHLH protein similar to *R* gene in maize, which also promotes anthocyanin biosynthesis (Goodrich *et al*., 1992). Though the *del* mutant affects only the corolla tube, the *Del* gene was shown to be active in corolla lobes (Almeida, 1989), where it is though to function redundantly with the related, lobe-specific bHLH gene *Mutabilis*, described later. *Del* had been mapped in family V1640 to the end of LG4 Feng *et al.*, (2009; Figure 34b). Using the existing genotype data for family V1640, regression analysis revealed no correlation between *Del* genotypes and pigmentation in either corolla lobe or tube (Figure 42a, b).



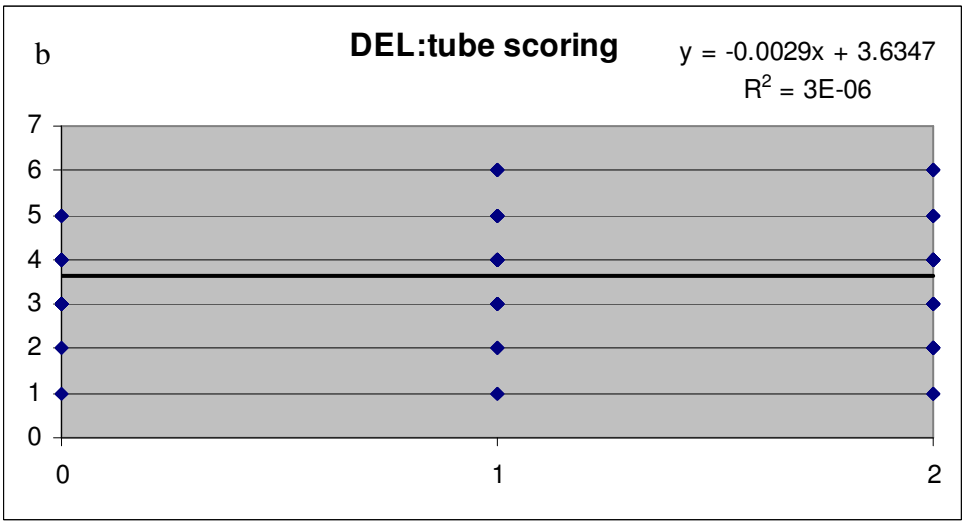


Figure 42. Simple linear regression analysis of lobe and tube colour and *Del* genotypes in F_2 family V1640 (n=83) The dose of the *A. majus* allele of *Del* is shown on the x-axis and the colour score, on a scale of 0-10 on the y-axis. (a) Corolla lobes, R^2=0.01, P=0.28. (b) Corolla tubes, R^2=3 x 10^{-6}, P=0.99).

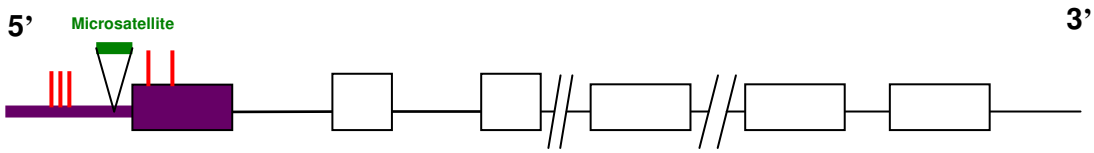
b) *Mutabilis* (*Mut*)

The *Mutabilis* (*Mut*) mutant was described by Stubbe (1966) and reduces corolla lobe pigmentation. The gene was isolated by Cathie Martin and colleagues in 2001. It was identified as encoding a bHLH protein similar to *Del* and is expressed in corolla lobes but not the corolla tube (Cathie Martin, personal communication). It is therefore likely to function redundantly with *Del* in lobes and can account for the loss of pigmentation only from the tube in *del* mutants. Sequences of genomic *Mut* clones were kindly provided by Cathie Martin. When combined, they suggested that the *Mut* gene consists of 6 exons and 5 introns (Figure 43a).

Partial *Mut* sequences were amplified from the 5' upstream region, the first exon, second exon and part of the second intron in an attempt to find allelic variation between species. The upstream region carried a (CT) microsatellite repeat that differed in length between the sequenced alleles. It was 34 bp long in *A. majus* JI.7 (*JI7Mut*) and *A. charidemi* (*AcMut*) and 36 bp long in *A. majus* JI.98 (*JI98Mut*; Figure 43b). Because both *A. majus* lines have fully pigmented flowers, this polymorphism is unlikely to reflect anthocyanin biosynthesis. Five additional nucleotide substitutions and an indel of 1 bp were also present.

In the region spanning the first and second introns, the two *A. majus* alleles, *JI7Mut* and *JI98Mut*, were identical and their sequence is therefore referred to *AmMut*. However, *AcMut* had an 11-bp deletion close to the start of the first intron, relative to *AmMut*. This polymorphism could potentially affect activity of the *Mut* allele, for example by disrupting splicing. Comparing the fully mature mRNA levels of *Mut* with qRT-PCR or Northern blots could test this hypothesis, though it was not carried out due to time limitations. The sequence alignments also revealed an additional 3bp indel and four nucleotide substitutions in the first intron (Fig. 44).

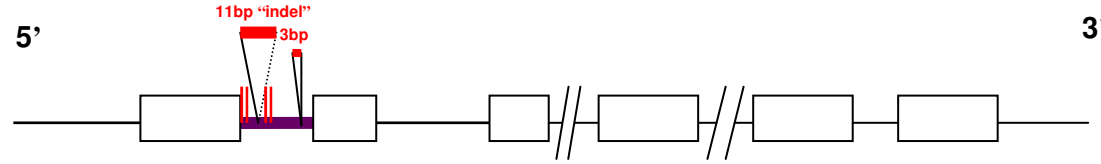
(a)



(b)

```
JI7Mut          GAGTCACAACATATGAAATTTAGGTAACAGAAAATAACCGTTCGGTGGGTACCAACCAGT 60
JI98Mut         GAGTCACAACATATGAAATTTAGGTAACAGAAAATAACCGTTCGGTGGGTACCAACCAGT 60
AcMut           GAGTCACAACATATGAAATTTAGGTAACAGAAAATAACCGTTCGGTGGGTACCAACCAGT 60

JI7Mut          CAAATCCCCACCACTCCCTCTTCTAATTAATTACATGACAACCAGAATTAATTCATTAAA 120
JI98Mut         CAAATCCCCACCACTCTCTCTTCTAATTAATTACATGACAACCAGAATTAATTCATTAAA 120
AcMut           CAAATCCCCACCACTCCCTCTTCTAATTAATTACATGACAACCAGAATTAAATCATTAAA 120

JI7Mut          ATTCTATCAATTCTCTTACAAATCCTACTTTCTTATCCTTCTAAAAGGCTAAAACAAATC 180
JI98Mut         ATTCTATCAATTCTCTTACAAATCCTACTTTCTTATCCTTCTAAAAGGCTAAAACAAATC 180
AcMut           ATTCTATCAATTCTCTTACAAATCCTACTTTCTTATCCTTCTAAAAGACTAAAACAAATC 180

JI7Mut          AAATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC--AGCCCTGAAAGCTAATCAA 238
JI98Mut         AAATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCAGCCCTGAAAGCTAATCAA 240
AcMut           AAATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC--AGCCCTGAAAGCTAATCAA 238

JI7Mut          CAACAAATCTTACATCTTGTACTTGTTTATACGTGTGTATGTATACTTTCTTGATTCTCT 298
JI98Mut         CAACAAATCTTACATCTTGTACTTGTTTATACGTGTGTATGTATACTTTCTTGATTCTCT 300
AcMut           CAACAAATCTTACATCTTGT-CTTGTTTATACGTGTGTATGTATACTTTCTTGACTCTCT 297

JI7Mut          TTTATGTCTTTTGTAGATGATAATTAGCATGTCACACTGGTATTTTCCTTTAGTCCTCTC 358
JI98Mut         TTTATGTCTTTTGTAGATGATAATTAGCATGTCACACTGGTATTTTCCTTTAGTCCTCTC 360
AcMut           TTTATGTCTTTTGTAGATGATAATTAGCATGTCACACTGGTATTTTCCTTTAATCCTCTC 357

JI7Mut          TCCCACCAAAATTTTCAGCTAAAATCGGAACATCATTGAGCGTTTCATCTACCTAAATTA 418
JI98Mut         TCCCACCAAAATTTTCAGCTAAAATCGGAACATCATTGAGCGTTTCATCTACCTAAATTA 420
AcMut           TCCCACCAAAATTTTCAGCTAAAATCGGAACATCATTGAGCGTTTCATCTACCTAAATTA 417

JI7Mut          TTCTCACTTGTTTTTTCAGCTGATTCAATAGGGCTTTATGTTCTTGATACCACTTTTTCA 478
JI98Mut         TTCTCACTTGTTTTTTCAGCTGATTCAATAGGGCTTTATGTTCTTGATACCACTTTTTCA 480
AcMut           TTCTCACTTGTTTTTTCAGCTGATTCAATAGGGCTTTATGTTCTTGATACCACTTTTTCA 477

JI7Mut          AGTCTCAATGAGTTTATAGAAATAAAGTAAGGGAGTATTTTCTCCTGATTAAAAGGCAGA 538
JI98Mut         AGTCTCAATGAGTTTATAGAAATAAAGTAAGGGAGTATTTTCTCCTGATTAAAAGGCAGA 540
AcMut           AGTCTCAATGAGTTTATAGAAATAAATTAAGGGAGTATTTTCTCCTGATTAAAAGGCAGA 537

JI7Mut          AAATAACCTGAAAAAAAGTTTGGAGTTTTGAGAAGTACTAAAATGGCTGCACCGCCGACA 598
JI98Mut         AAATAACCTGAAAAAAAGTTTGGAGTTTTGAGAAGTACTAAAATGGCTGCACCGCCGACA 600
AcMut           AAATAACCTGAAAAAAAGTTTGGAGTTTTGAGAAGTACTAAAATGGCTGCACCGCCGACA 597

JI7Mut          AAGAGTCGGCTTCAGTGCATGTTGCAGGCGGCGGTTCAAGCTGTTCAGTGGACTTACACC 658
JI98Mut         AAGAGTCGGCTTCAGTGCATGTTGCAGGCGGCGGTTCAAGCTGTTCAGTGGACTTACACC 660
AcMut           AAGAGTCGGCTTCAGTGCATGTTGCAGGCGGCGGTTCAAGCTGTTCAGTGGACTTACACC 657

JI7Mut          CTCTTCTGGCAACCGTGCCCCCAACAAGGGTAAACATTAGAAGAATTAAATCATTTCTTA 718
JI98Mut         CTCTTCTGGCAACCGTGCCCCCAACAAGGGTAAACATTAGAAGAATTAAATCATTTCTTA 720
AcMut           CTCTTCTGGCAACCGTGCCCCCAACAAGGGTAAACATTAGAAGAATTAAATCATTTCTTA 717

JI7Mut          ATACTCTCA 727
JI98Mut         ATACTCTCA 729
AcMut           ATACTCTCA 726
```

Figure 43. Allelic variation in and around the first exon of *Mut* Sequences are shown for alleles from *A. majus* line JI.7 (*JI7Mut*) and line JI.98 (*JI98Mut*) and from *A. charidemi* (*AcMut*). (a) Schematic diagram shows the structure of *Mut* locus, which contains six exons and five introns. The sequenced region in shown in purple and polymorphisms are marked in red. (b) Sequence comparisions. Nucleotides of the first exon are shown in green, the variable microsatellite region in red and invariant nucleotides are shaded.

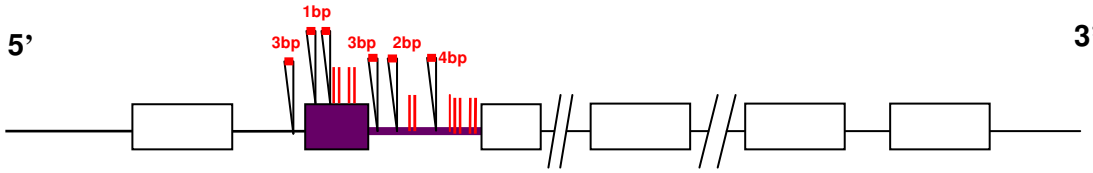



```
AmMut           TTCCGTGCTCACTTCTGTTGGTTGCACTGTTTTTCTTGTTTTTATTACTCCGTTGTAATA 60
AcMut           TTCCGTGCTCACTTCTGTTGGTTGCACTGTTTTTCTTGTTTTTATTACTCCGTCGTAATA 60

AmMut           TCCATCACGCCACACTAGAATTCTTCACATCAAACATATATGTACATGGGATATCAAATG 120
AcMut           TCCATCACGCCACACTAGAATTCTTCACATCAAACATATATGTACATGGAATATCAAATG 120

AmMut           TTGCATTGAAGTTATATATCTAAGGATATATATCGAAAATATGATTATCTACTCATTTAT 180
AcMut           TTGCATTGAAGTT-----------GATATATATCGAAAATATGATTATCTACTCATTTAT 169

AmMut           GTTGAACAATAAATATAAAATAATCTAAACACTTATGTTTTGAGATTAGACAACTACTCA 240
AcMut           GTTGAACAATATATATAAAATAATCTAAACACTTATGTTTTGAGATTAGACAACTACTCA 229

AmMut           ATATGATATTAGAATCAGATAAAAAAAAGTCTGGTGCGAATTCTACCAACACTACTTATC 300
AcMut           ATATGATATTAGAATCAGAAAAAAAAAGTCTGGTGCGAATTCTACCAACACTACTTATC 289

AmMut           AAAAAAATGTCGCCGGTAAAGTGCACTGAAAAAAAGGCAAATTATATTCCGTAACTAATG 360
AcMut           AAAAAAATGTCGCCGGTAAAGTGCACTGAAAAAAAGGCAAATTATATTCCGTAACTAATG 349

AmMut           ATGCGTACGTTAAGAAATATATAATATAAAAATAATGAATATTCATTATAACTTGAGCTT 420
AcMut           ATGCGTACGTTAAGAAATATATAATATAAAAATAATGAATATTCATTATAACTTGAGCTT 409

AmMut           TTTAGATGATCAGTGGCATTTCCGTAGTTCAAAACACACTTGAAGATGTTAATTAAAGAA 480
AcMut           TTTAGATGA---GTGGCATTTCCGTAGTTCAAAACACACTTGAAGATGTTAATTAAAGAA 466

AmMut           AATGATTTAATTCTTCTAATGTTTACCCTTGTTGGGGGCACGGTTGCCAGAAGAGGCTGT 540
AcMut           AATGATTTAATTCTTCTAATGTTTACCCTTGTTGGGGGCACGGTTGCCAGAAGAGGCTGT 526

AmMut           AAGTCCACTGAACAGCTTGAACCGCCGCCTGCAACATGCACTGAAGCCGACTCTTA 596
AcMut           AAGTCCACTGAACAGCTTGAACCGCCGCCTGCAACATGCACTGAAGCCGACTCTTA 582
```

Figure 44. Allelic variation in the first intron of *Mut* Labelling is the same as for Fig. 43.

More variation between *Mut* alleles was detected in the second exon than the first. Two 1 bp indels and four nucleotide substitutions distinguished *AcMut* from *AmMut* (Fig. 45). The two 1 bp indels lead to frame-shift mutations. However, the amino acid sequence necessary for function of Mut is not known, so the possibility still remains that this part of the protein is not necessary, for example because it is cleaved from the active protein during post-translational processing.

The second intron also varied between allele, with three indels and six nucleotide substitutions (Figure 45).




```
AmMut           AACAGGAAGGTGAGCACGGAAGAGGCCTCGCTCCAAAGAAGCCAACAACTAAAAGAATTA 60
AcMut           AACAGGAAGGTGAGCACGGAAGAGGCCTCGCTCCAAAGAAGCCAACAACTAAAAGAATTA 60

AmMut           TACGATTCCC-TCGCTGCCGGAG-AATCAAACAGCCAGGCGCAGCAGCGGCGGCCCTCCG 118
AcMut           TACGATTCCCGTCGCTGCCGGAGTAATCAAACAGCCAGGCGCAGTTGCGGCGGCCCTCCG 120

AmMut           CCTCGTTGTCACCGGAAGATCTAACTGAGTCTGAGTGGTATTATCTAATGTGTGTTTCCT 178
AcMut           CCTCGTTGTCACCGGAAGATCTAACTGAGTCTGAGTGGTATTATCTAATGGGTGTATCCT 180

AmMut           TCTCTTTTCCTCCGGGCCTCGGGTAAGAATTCTTAAAAGCATATATACTCTTTTACATTA 238
AcMut           TCACTTCTCCTCCGGGCCTCGGGTAAGAATTCTTAAAAGCATATATACTCTTTTACATTA 240

AmMut           AATTCTAATCAGACAATCGCATGCTCTAATGATAAAAAGTTTCATTTTCGATATAT--GT 296
AcMut           AATTCTAATCAGACAATCGCATGCTCTAATGA---AAAGTTTCATTTTCGATATATACGT 297

AmMut           TTTAATATTTAATTGGAACAATTTTTAAAAAACGTGAGTCATGGCATAGAGCCATAGACT 356
AcMut           TTTAATATTTAATTGGAACAATTTTTAAAAAACTTGAGTCATGGCATAGAGCCATAGACT 357

AmMut           AAATCATTGTCCGTGTTCCATTTCCTGTATCAATATCTATTTGATTGAGTGAAGTGTTTC 416
AcMut           AAATCATTGTCCGTGTTCCATTTCTTGTATCAATATCTATTTGATTGAGTGAAGTGTTTC 417

AmMut           CGAAGTTTGACATTAATTCTATTCAGCGATGGTAAGATATTTATTAAGATGAGTATTAAA 476
AcMut           CGAAGTTTGACATTAATTCTATTCAGCGATGGTAAGATATTTATTAAGATGAGTATTAAA 477

AmMut           TTTCAACCTGCACCAATTTTGGGACCATGAAATTTGTAATTATGTGTACCTTGAGTCTTA 536
AcMut           TTTCAACCTGCACCAATTTTGGGACCATGAAATTTGTAATTATGTGTACCTTGAGTCTTA 537

AmMut           AAGTCCGTGGGGAGTGGTTGCGTCTGTAAAATGAAATATGAAATAAAAAGAAAGGCAAAC 596
AcMut           AAGTCCGTGGGGAGTGGTTGCGTCTGTAAAATGAAATATGAAATAAAAAGAAAGGCAAAC 597

AmMut           TGTAATAAGTTCATAACCATGCATGGAAATTTTCTTTTCCATTTCTGCTAAAAATTTAAG 656
AcMut           TGTAATAAGTTCATAACCATG----GAAATTTTCTTTTCCATTTCTGCTAAAATTTTAAG 653

AmMut           TGAGCAATTAAGATTATAAGCAATTAAGAAATTTCTTATAATACTAAGATTAGTACTTTT 716
AcMut           TAAGCAATTAAGATTATAAGCAATTAAGAAATTTCTTATAATACTAAGATCAGTACTTTT 713

AmMut           TTCTGATTCAGAGAATTCTAACTCTTGCTTCCAGATTAATGTTAGGACAGTGGATATTTT 776
AcMut           TTCTGATTCAGAGAATTCTAACTCTTGCTTCCAGATTAATGTTAGGACAGTGGATATTTT 773

AmMut           TTTAGTATGCATGCTTACTGCGAAGTGTGCAACATATATACATGTTATCTTTAAATTCCT 836
AcMut           TTTAGTATGCATGCTTACAGCGAAGTGTGCAACATATATACATGTTATCTTTAAATTCCT 833

AmMut           TTTTTTAGTTAATCACTCGGTTCATTTCAATTAGATATCTACATATGATTTATATTGAAG 896
AcMut           TTTTTTAGTTAATCACTCGGTTCATTTCAATTAGATATCTACATATGATTTATATTGAAG 893

AmMut           TTGTGCATTATTATCTTCATAATTATAAGTAACATACAAATCCTTACGGCTTTACAAAAT 956
AcMut           TTGTGCATTATTATCTTCATAATTATAAGTAACATACAAATCCTTACGGCTTTACAAAAT 953

AmMut           AGGGATTAATTTGAACAATTAATCTGTAACTTAACTTAAGATATTGAATTATTTGTTCCC 1016
AcMut           AGGGATTAATTTGGACAATTAATTTGTAACTTAACTTAAGATATTGAATTATTTGTTCCC 1013

AmMut           CCACTAATATCAACAGAGTAATAATTGAATTTAGCGCGTAATGAACTTTGAAAAGATTTG 1076
AcMut           CCACTAATATCAACAGAGTAATAATTGAATTTAGCGCGTAATGAACTTTGAAAAGATTTG 1073

AmMut           ATTTTGTGAATGTGACATTTCTTATCTTATAATATAATTTGAGCAATCTAGATAAAAAAT 1136
AcMut           ATTTTGTGAATGTGACATTTCTTATCTTATAATATAATTTGAGCAATCTAGATAAAAAAT 1133

AmMut           ATGTTCGCATAACACT 1152
AcMut           ATGTTCGCATAACACT 1149
```

Figure 45. Allelic variation in the second exon and second intron of *Mut* Labelling is the same as for Figure 43.

The variation found in *Mut*, in both coding and non-coding regions, was higher than seen for other loci. One explanation is that selection on *Mut* has been relaxed because it functions redundantly with *Del*.

The polymorphism in the first intron was used to genotype members of the F_2 family V1640. This allowed *Mut* to be mapped to 56 cM in LG4 (Fig. 46).

F_2 genotypes were also used to test for a correlation with flower colour phenotypes using simple regression and one-way *ANOVA*. No significant correlation was detected (Fig. 47 a,b).

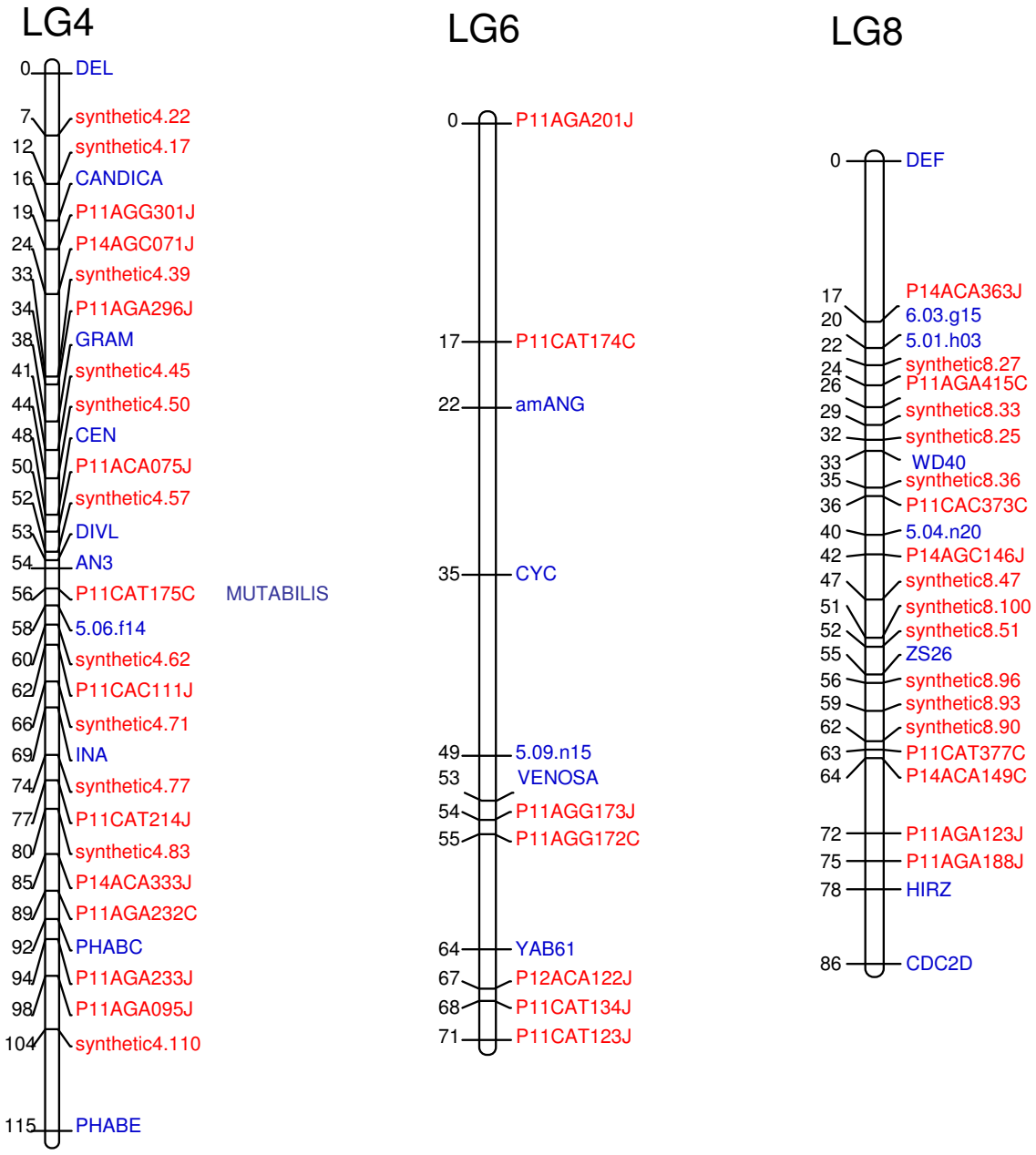


Figure 46. Map positions of *Mut*, *Ve* and a *WD40* Mapping was carried out in F_2 family V1640, using JoinMap 3.0 software**.**

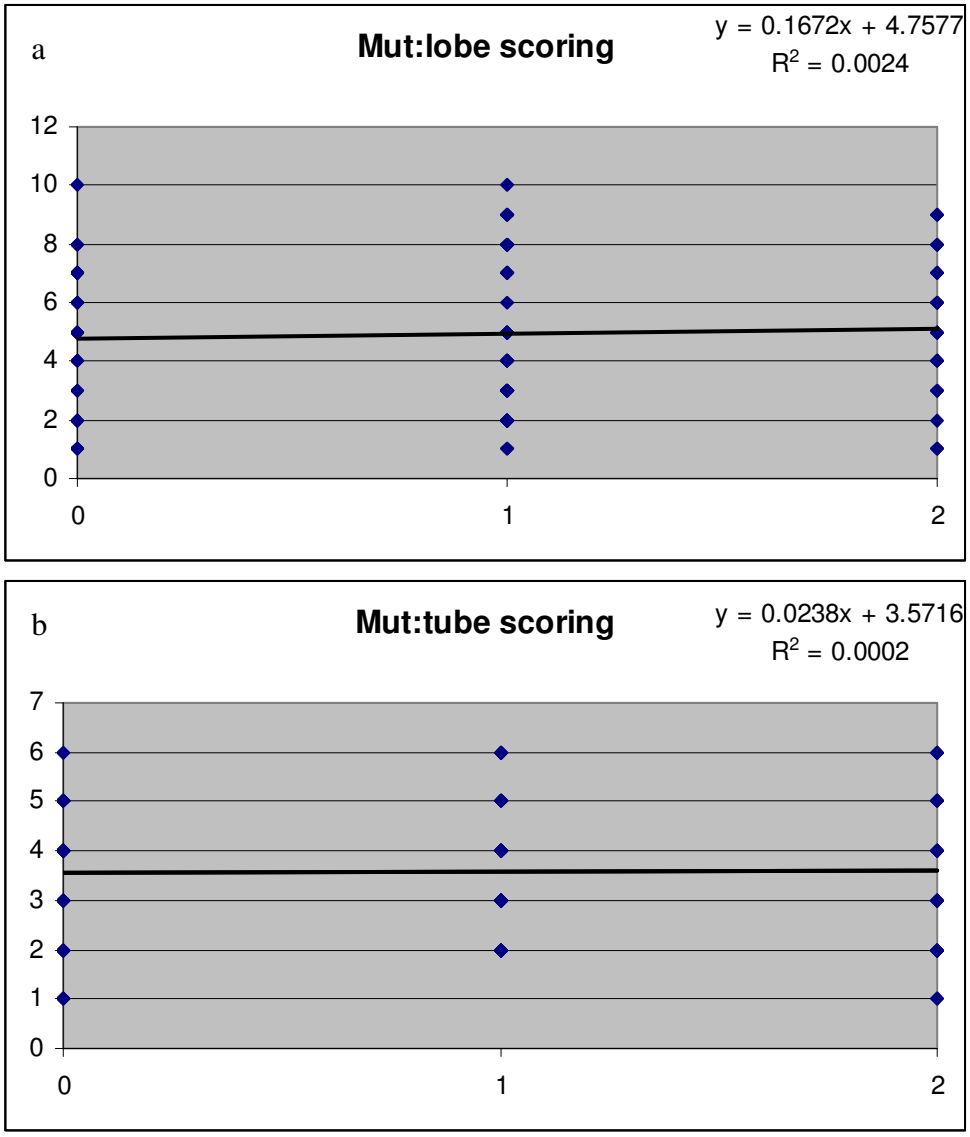


Figure 47. Simple linear regression of flower colour and *Mut* phenotypes in family V1640 (n=110) The dose of the *A. majus* allele of *Mut* is shown on the x-axis and colour scoreson the y-axis. For corolla lobe colour (a) R^2=0.002, P=0.61 for corolla tube colour (b) R^2=0.0002, P=0.89.

ii. *MYB* regulatory genes

a) *Rosea1* (*Ros1*) and *Rosea2* (*Ros2*)

Ros1 and *Ros2* belong to a small family of R2R3 MYB proteins which also contains *Venosa* (*Ve*). *Ros1* and *Ros2* represent a recent tandem duplication at the *Ros* locus, identified from mutations that alter flower pigmentation (Schwinn *et al.*, 2006). Analysis of mutants suggests that *Ros1* and *Ros2* control both the intensity and pattern of magenta colour in flowers by regulating expression of ABP structural genes. Though *Ros1* and *Ros2* encode very similar proteins (84% amino acid identify), these have been reported to show different interactions with the bHLH proteins *Del* and *Mut* (Schwinn *et al*., 2006). *Ros1* interacts with *Del* in the corolla tube but interacts specifically with *Mut* in lobes, even though *Del* is also expressed here. In contrast, *Ros2* interacts only with *Del*, in both lobes and tubes (Schwinn *et al*., 2006). There has also been a suggestion that *Ros2* might act as repressor, because the *Ros2* gene is disrupted in *A. majus*, which has fully pigmented flowers, but expressed in other species with paler flowers. However, *Ros2* is able to activate anthocyanin biosynthesis when expressed from the *35S* promoter in *ros1* mutant petal cells, which argues against its role as a repressor (Schwinn *et al*., 2006).

In the previous chapter, *Ros1* genotypes had been found to show a significant correlation with flower colour in the V1640 family, particularly in petal tubes. *Ros2* had not been tested separately because the two loci are very closely linked (<5 kb apart) (Schwinn *et al*., 2006). In this chapter, I tested the correlation of *Ros1* genotype and flower colour in two NILs families, M241 and M244.

The M241 family, which segregated a region of LG3, showed a significant correlation between *Ros1* genotype and either lobe or tube colour (Figure 48a, b). However, in linear regression, *Ros1* genotypes explained only ~9% of the petal lobe colour variation in M241, compared to 26% in the F_2 population V1640. This could be explained by segregation of a second, linked locus also affecting lobe colour in the F_2 which was not

segregating in the NIL. Alternatively, the effect of *Ros* might be dependent on the alleles of other genes that were present in the F_2 but not the NIL. For corolla tubes, *Ros* genotype could explain a similar proportion of the variance in the F_2 (50%) and the NIL (49%). This suggests that the effect of *Ros* in the corolla tube is not dependent on alleles that were lost during generation of M241 and favours the presence of a second linked locus with a major effect on petal lobe colour in the F_2 but not the NIL.

A similar result was found in M244 family, which had been found to segregate parts of both LG3 and LG7. If this family was heterozygous at the two major-effect QTL detected in the F2, it was expected to show similar variation in flower colour to V1640. However, although flower colour and *Ros1* genotype were still significantly correlated (P=0.02) *Ros1* genotype could only explain about 6% of petal lobe variation in M244, though the correlation between corolla tube colour and *Ros1* genotype remained almost unchanged (Fig. 49 a, b). This again argues for a second gene linked to *Ros* which has a major effect on petal lobe colour.

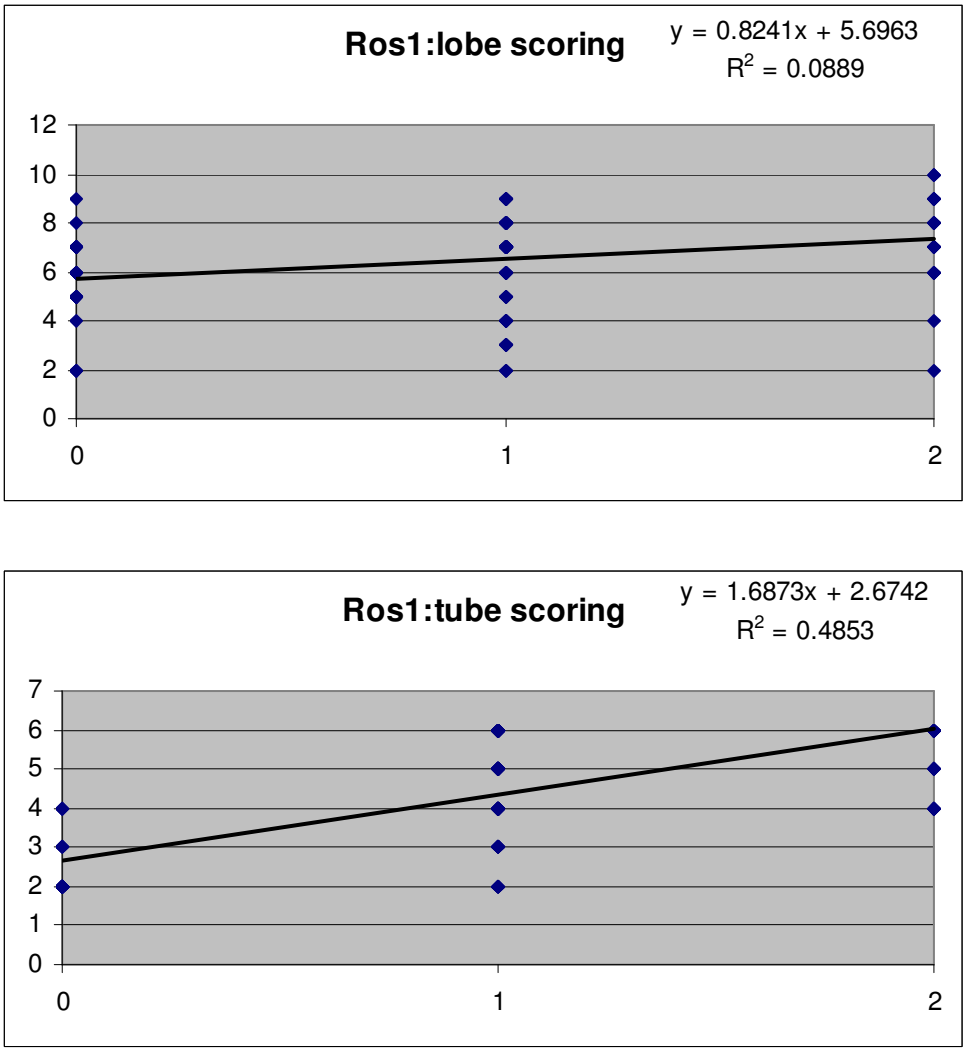


Figure 48. Simple linear regression of flower colour and *Ros1* genotype in NIL family M241 (n=109) Dose of the *A. majus Ros1* allele is shown on the x-axis and colour score, on a scale of 0-10 on the y-axis. For corolla lobes (a) R^2 =0.09, P= 0.0016 and for tubes (b) R^2=0.49, P=4.26 x 10^{-16}.

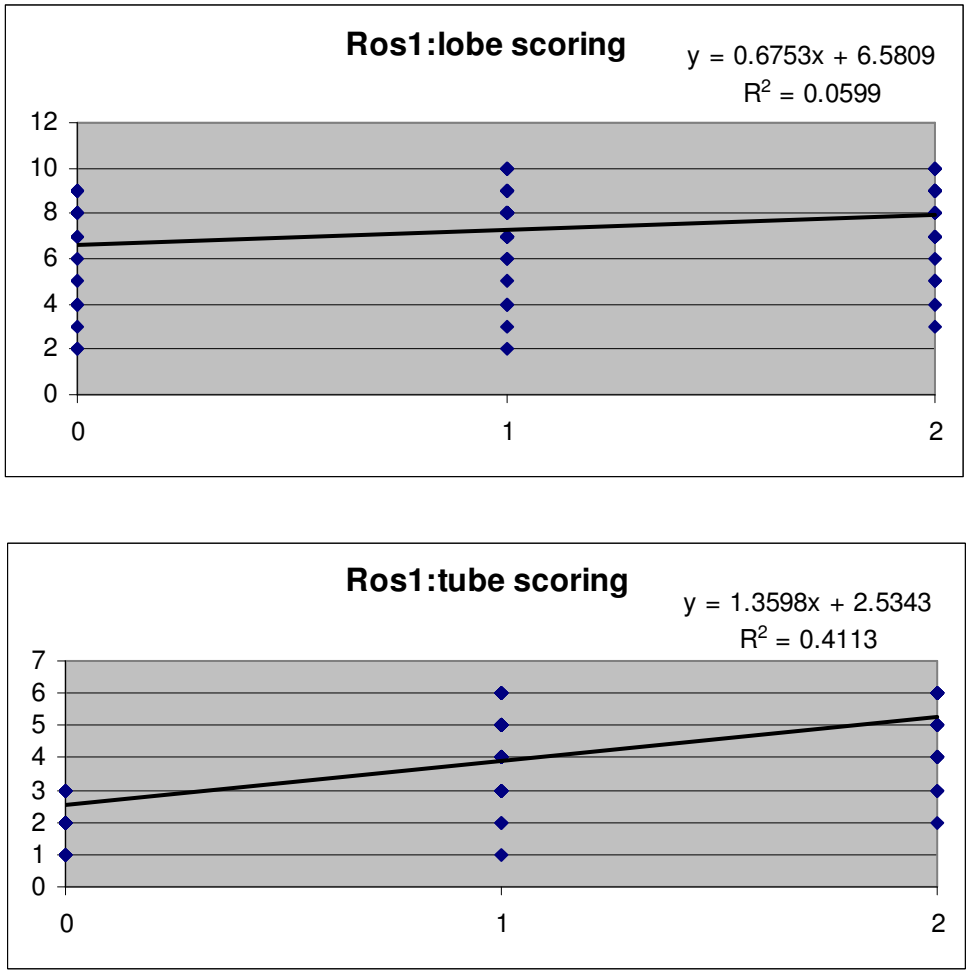


Figure 49. Simple linear regression analysis between flower colour and *Ros1* in the NIL family M244 (n=87) Axes are shown as in Figure 50. For corolla lobes (a) R^2=0.06, P=0.0215 and for tubes R^2=0.411, P=1.38 x 10^{-8}.

b) *Venosa* (*Ve*)

Venosa (*Ve*) was originally identified as a locus affecting the pattern of anthocyanin pigmentation (Stubbe, 1966). In a genetic background with reduced *Ros* activity, the dominant *Ve* allele promotes anthocyanin production in dorsal corolla epidermis cells over veins. This pattern is seen in several *Antirrhinum* species, including *A. molle*, suggesting that the dominant *Ve* allele originally characterised as a mutation in *A. majus* might have been introgressed from another species. The *Ve* locus was cloned and its structure characterized recently (Schwinn *et al*., 2006). It encodes a Myb protein closely related to *Ros1* and *Ros2* (77% identify in cDNA sequences), though the proteins are

more similar in their Myb-domains than C-terminal regions (Schwinn *et al*., 2006; Shang, 2006). The gene contains three exons and two introns. *In situ* hybridization, using the C-terminal domain sequence as a probe, showed that *Ve* is transcribed in epidermal and sub-epidermal cells over veins, even though it promotes anthocyanin synthesis only in epidermal cells (Shang *et al*., 2011). This suggests that other genes that are expressed only in epidermal cells are needed to activate anthocyanin synthesis (e.g., *bHLH* genes). *Ve* can also promote anthocyanin biosynthesis in cells away from the vein when it is misexpressed (Schwinn *et al*., 2006; Shang *et al*., 2011).

To find allelic variation in *Ve* between *A. majus* (*JI7Ve*) and *A. charidemi* (*AcVe*) regions of all exons and the third intron were sequenced. The first intron, first exon and second exon were found to be identical between the species, but the last exon differed by nine nucleotide substitutions and a 2-bp indel (Fig. 50a). The first SNPs in this exon caused an amino acid substitution at residue 120 from Asparagine (N) in *A. majus* (in both the published cDNA sequence and *AcJI7*) to Aspartic Acid (D) in *A. charidemi*. However it is not clear whether the published sequence, from an active *Ve* allele, had come from a species other than *A. majus*. The second and the third SNPs (which are common to the published sequence and *AcVe*) are synonymous (Fig. 50a, b). The fifth, sixth and seventh SNPs were shared by both *A. majus* alleles, but caused an amino acid substitutions from Arginine (R) in *A. majus* to Glutamine (Q) in *A. charidemi* at position 154; from Aspartic Acid (D) to Asparagine (N) at position 160 and from Glutamine (Q) to Glutamic acid (E) at position 163. The eighth SNP distinguished *JI7Ve* from the other two alleles causing an amino acid substitution at position 194 (Figure 50b). The ninth SNP also distinguished *JI7Ve* from the other two alleles, but was synonymous, while both *AcVe* and *JI7Ve* differed from the published cDNA sequence immediately before the stop codon. Because these alleles carry a disrupted stop codon they potentially encoded a Ve protein that is longer than the published amino acid sequence (GI:82570708). Unfortunately, 3'sequences did not extend far enough to determine the potential C-terminal amino acid sequences of the novel proteins.

```
cDNA            AAAAACTTCTGGAATACCCACTTCGAGAAGAAGTCCGGAGAACGAGAGAATACGGAAAAT 360
JIVe            ----------------------TTCGAGAAGAAGTCCGGAGAACGAGAGAATACGGAAAAT 39
AcVe            ----------------------TTCGAGAAGAAGTCCGGAGAACGAGAGAATACGGAAGAT 39

cDNA            ATAAACCCGAAACTCATCAACTCGAGCAATATAATAAAACCCCAACCTCGTACCTTCTTG 420
JIVe            ATAAACCCGAAACTCATCAACTCCAGCAATATAATAAAACCCCAACCTCGTACCTTCTTG 99
AcVe            ATAAACCCGAAACTCATCAACTCGAGCAATATAATAAAACCCCAACCTCGTACCTTCTTG 99

cDNA            AAACTGCGTCCCAAGGAAACAAAGAAACAGAAAAATATACGGAACGTTTGTACAGCAAAT 480
JIVe            AAACTGCGTCCTAAGGGAACAAAGAAACAGAAAAATATACGGAACGTTTGTACAGCAAAT 159
AcVe            AAACTGCGTCCCAAGGAAACAAAGAAACAGAAAAATATACAGAACGTTTGTACAGCAAAT 159

cDNA            GATGACAAACAGCAGCCGTTGTCCACGTCCGGACAGTTAGAAGAAGTGAATGAACGCATT 540
JIVe            GATGACAAACAGCAGCCGTTGTCCACGTCCGGACAGTTAGAAGAAGTGAATGAACGCATT 219
AcVe            AATGACAAAGAGCAGCCGTTGTCCACGTCCGGACAGTTAGAAGAAGTGAATGAACGCATT 219

cDNA            CGGTGGTGGAGTGAATTGCTCGATTTTGCGGACTACGTAGATTAA 585
JIVe            CGGTGGTGGAGTGAATTGCTCGATTTTGCGGACTACGTAAAA--- 261
AcVe            CGGTGGTGGAGTGAATTGCTCGATTTTGCGGACTACGTAGAA--- 261
```

(b)

```
                IKRGSFTRDEVDLIVRLHKLLGNRWSLIAGRLPGRTGNDVKNFWNTHFEKKSGERENTEN 120
JIVe            -------------------------------------------------FEKKSGERENTEN 13
AcVe            -------------------------------------------------FEKKSGERENTED 13

full            INPKLINSSNIIKPQPRTFLKLRPKETKKQKNIRNVCTANDDKQQPLSTSGQLEEVNERI 180
JIVe            INPKLINSSNIIKPQPRTFLKLRPKGTKKQKNIRNVCTANDDKQQPLSTSGQLEEVNERI 73
AcVe            INPKLINSSNIIKPQPRTFLKLRPKETKKQKNIQNVCTANNDKEQPLSTSGQLEEVNERI 73

full            RWWSELLDFADYVD* 195
JIVe            RWWSELLDFADYVKX 87
AcVe            RWWSELLDFADYVEX 87
```

Figure 50. Allelic variation at *Ve* The DNA sequence of the final exon in JI.7 (*JIVe*) *A. charidemi* (*AcVe*) and of the published cDNA sequence, from a *Ve* allele of unknown origin in *A. majus* (GI:82570708) are compared in (a). The published cDNA sequence is shown in blue and invariant nucleotides are shaded. (b) Shows the amino acid translations from the nucleotide sequences, X = unidentified amino acid.

It is not clear whether the amino acid substitutions in this region of the Ve protein might affect its activity.

Although all characterized active *Ve* alleles appear to be expressed only over veins (Schwinn *et al*., 2006; Shang *et al*. 2006), it is possible that other *Ve* alleles might be more widely expressed and therefore contribute to differences in the general pattern or level of floral pigmentation between species. Therefore *Ve* was tested as a candidate for colour differences between *A. majus* and *A. charidemi*.

Ve was mapped in the V1640 F_2 population to 53 cM in LG6 using a dominant *Dde*I CAPS marker able to distinguish presence of the allele from *A. majus* (Fig. 46). No significant flower colour QTL had been detected in this chromosome. The correlation of

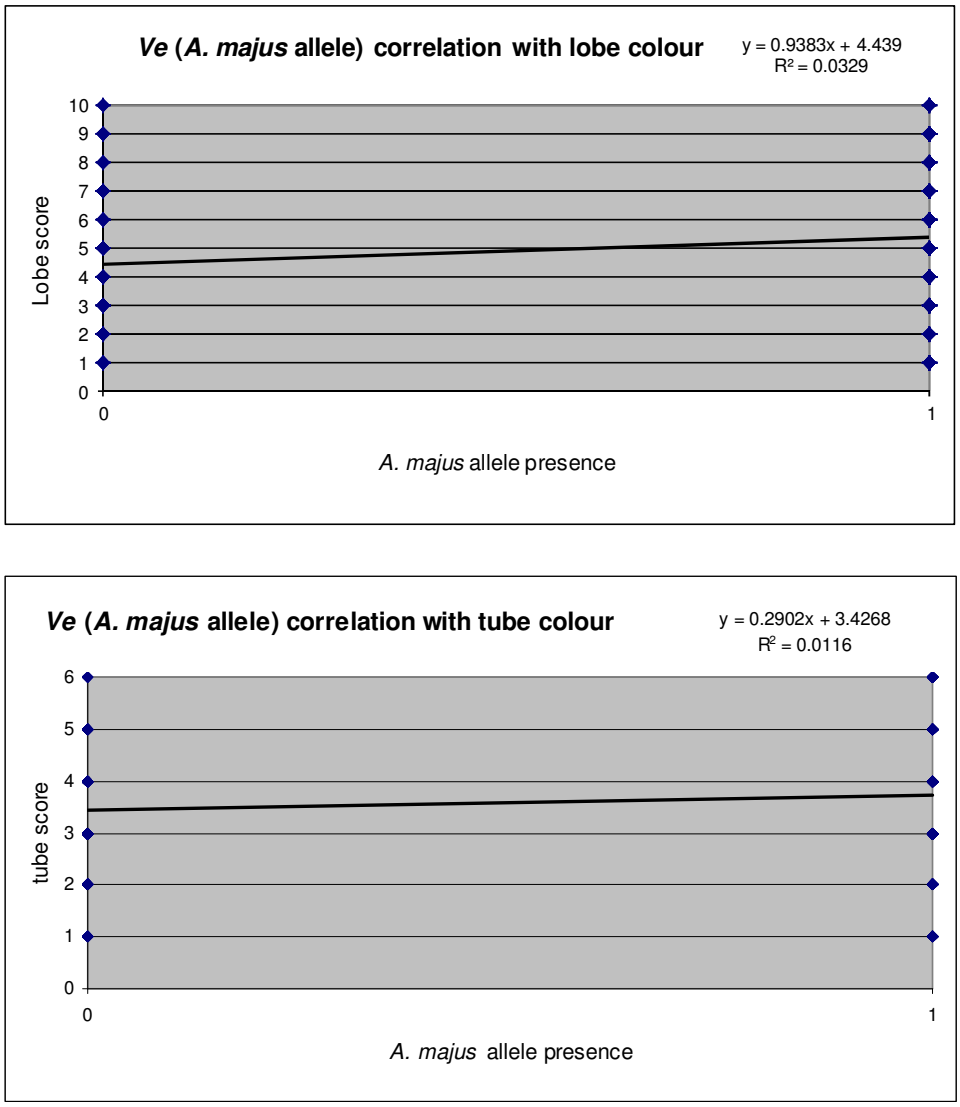


Figure 51. Simple linear regression of flower colour and *AmVe* in F_2 V1640 family (n=187) Dose of the *AmVe* allele is shown on the x-axis and colour score, on a scale of 0-10 on the y-axis. For corolla lobes (a) R^2= 0.033, P-value = 0.012 and for tubes (b) R^2= 0.012, P-value = 0.1412.

AmVe allele and lobe colour variation in V1640 family was found to be marginal (R^2= 0.033, P-value = 0.012 [Figure 51a]) but no correlation with tube variation (R^2= 0.012, P-value = 0.1412 [Figure 51b]).

iii. A *WD40* regulatory gene

WD40 proteins contain 4-16 copies of highly conserved repeating units, of ~40 amino acids, which usually end with Tryptophan and Aspartic acid (WD). WD40 proteins are found in all eukaryotes and are implicated in a wide variety of functions, including RNA processing, signal transduction, cytoskeleton assembly and cell division (reviewed by Smith *et al*., 1999). They form circular β-propeller structures in which each of the repeats contributes a blade of the propeller and act by binding to other proteins and promoting protein-protein interactions. A WD40 protein is required for assembly of transcription factor complexes containing bHLH and MYB proteins, including the complexes that regulate expression level of anthocyanin biosynthesis genes (Broun 2005; Ramsay and Glover 2005, Pathirana, 2009). The first WD40 found to be involved in control of anthocyanin production was *an11* in Petunia (de Vetten *et al*., 1997).

A likely orthologues of *an11* had been identified in *A. majus* and its cDNA sequences were kindly provided by Cathie Martin and Kathy Schwinn. Blasting the *A. majus* genome sequence showed that the coding region of the gene (called *WD40* here) contained no introns. Primers were designed from the cDNA sequence and used to amplify the alleles from *A. majus* lines, JI7 and JI98 (*JI7WD40* and *JI98WD40*) and from *A. charidemi* (*AcWD40*). The allele sequences differed by seven single-nucleotide substitutions (Figure 54). Only the first of these SNPs (at position 405) is within the coding region, which ends at position 686. This SNP causes an amino acid substitution of Threonine (T) in *A. majus* with Alanine (A) in *A. charidemi*.

Using a co-dominant *Msp* I marker to distinguish the *JI7WD40* and *AcWD40* alleles, I was able to map the locus in the V1640 F_2 population to 33cM in LG8 (Fig. 46). Although no significant QTL for flower colour had been detected at this position in the F_2 population, tube colour showed a marginally significant correlation with *WD40* genotype in linear regression (R^2=0.040, P= 0.172; Figure 53) with the *A. majus* allele decreasing pigmentation.

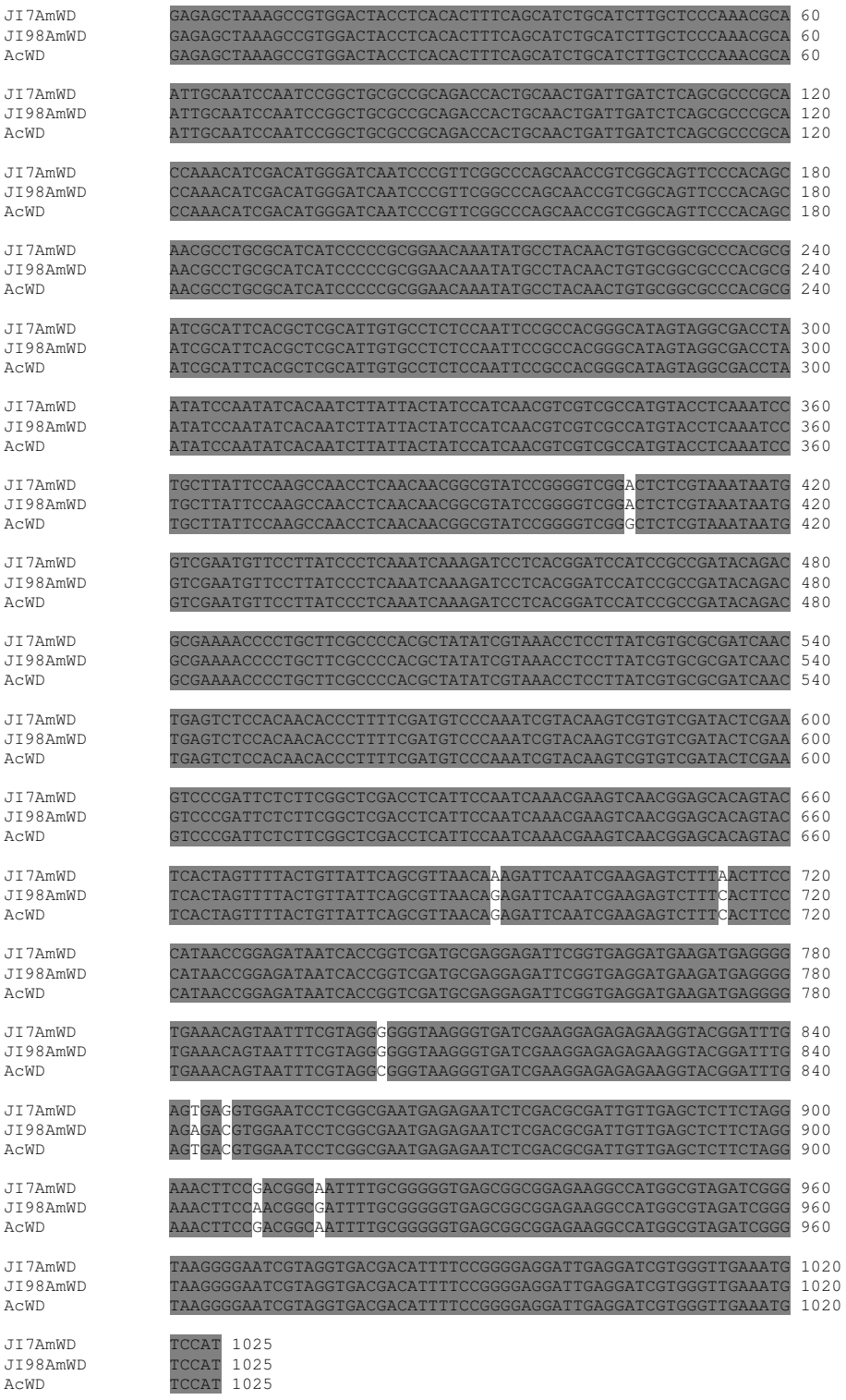


Figure 52. Allelic variation in *WD40*

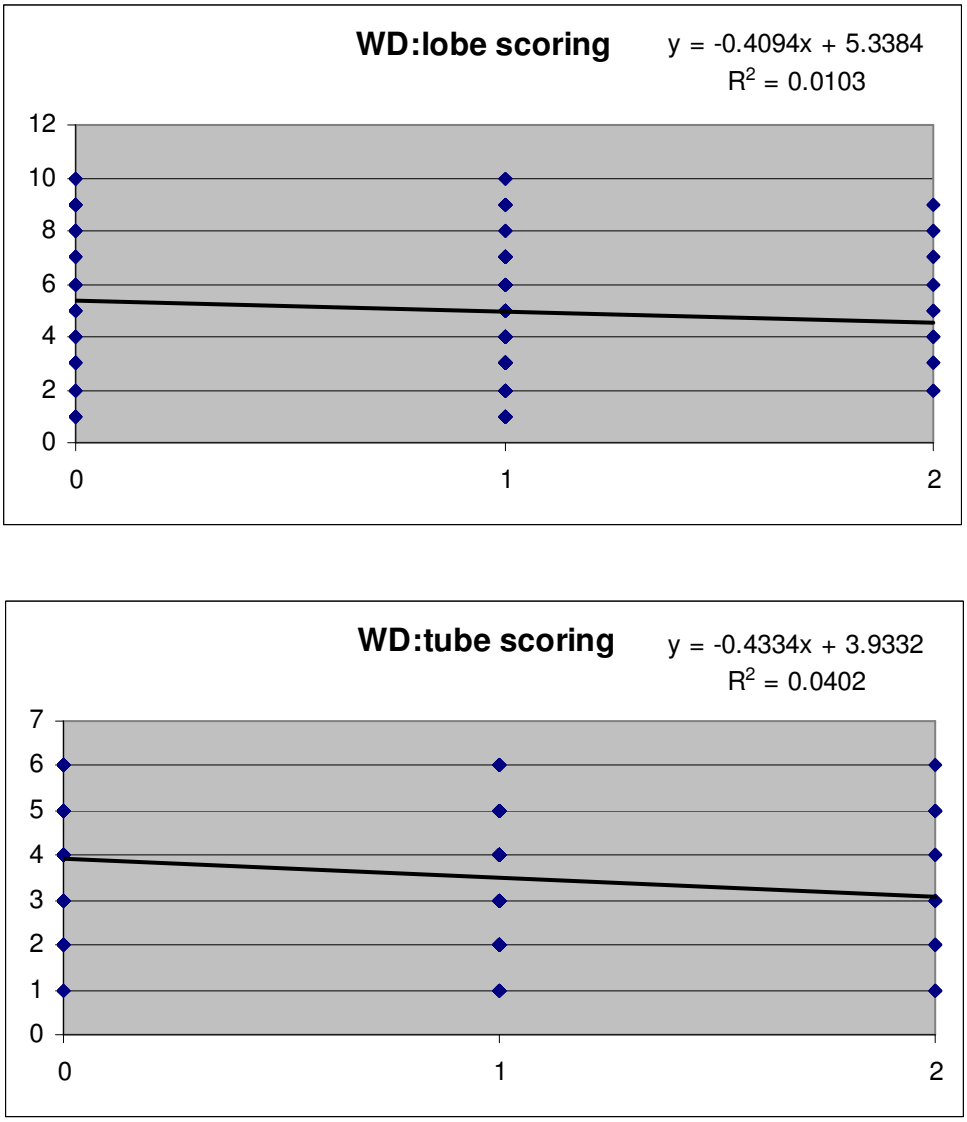


Figure 53. Simple linear regression of flower colour with *WD40* genotype in the V1640 F_2 population (n=140) The dose of the *A. majus* allele is plotted on the x-axis and colour scores on the y-axis. For lobe colour (a) R^2 = 0.00103, P= 0.231 and for tube colour (b) R^2=0.040, P=0.017.

4.3 Conclusion and Discussion

Genes involved in anthocyanin pigmentation showed allelic variation between and within species. Statistical tests revealed some significant correlations between flower colour and genotypes. This included the correlation of lobe colour with *Inc* genotype in the F2 family and in the NIL family M244. However, the NIL family M239, showed similar variation in flower colour to M244 but did not segregate at *Inc*, suggesting that the major QTL in LG7 is not *Inc*, but linked to it.

Analysis in the NILs families segregating at *Ros* detected a correlation between *Ros1* genotype and both corolla tube and lobe pigmentation as had been found in the F2 population. Therefore *Ros* remains a candidate for this QTL. However, in the NIL family M241, *Ros1* genotype explained only 9% of the variation in petal lobe variation, compared to 33% in the F2 population. In contrast, it explained a similar proportion of the variation in lobe colour in the F2 (60%) and in M241 (50%). One explanation for this difference is that a second QTL in LG3 has a major effect on lobe colour and that this QTL did not segregate in the NILs families. Alternatively, *Ros* might have a major effect in lobes that is dependent on alleles of other genes that were present in the F2 but not in the NILs families.

One candidate for a second QTL in LG3 is *Pal*. Because *Pal* is linked to *Ros*, it was difficult to separate the potential effects of the two genes in the F_2 family V1640. Although allelic variation was found for *Pal* in both coding and non-coding regions, *Pal* was not heterozygous in any of the NILs families, including M239, which was homozygous for the allele from *A. charidemi* (*AcPal*). This result clearly showed that *Pal* did not have involvement in family M239 corolla colour variation. Because *Ros* genotype explained a lot less petal lobe variation that in the F_2, this is consistent with involvement of a second QTL that could be *Pal*.

Candi was found to have a small but significant correlation with tube colour variation in V1640. Regression analysis suggests that *Candi* can explain about 6% of variation in the population. A small-effect QTL for tube colour detected at 12 cM LG4 can explain about 12 % of the variation found in V1640, very close to the location of *Candi*, which were reported at 16cM. *Del* and *Mut*, which were located in the same chromosome, did not show any correlation with flower lobe and tube variation.

Minor effect QTL affecting flower colour had been mapped to LG2 at 62cM and 88cM for lobe and tube variation, respectively. So far, the only gene involved in anthocyanin pigmentation reported to be located in this chromosome is *UFGT*, at 11cM in an *A. majus* x *A. molle* F_2 population (Schwarz-Sommer *et al*., 2003) making it remain possible that *UFGT* is a QTL in LG2. .

WD40 was mapped to 33cM in LG8, close to one of the small effect QTL that can explain some of the lobe colour variation mapping around 56cM. However, when regression analysis was performed with the *WD40* genotype in V1640, a significant correlation was found only for tube colour variation. This result was unexpected because in QTL analysis of tube colour variation, no significant QTL was found in LG8 when $\alpha = 0.05$. Nevertheless, a small but insignificant QTL peak had been found at 31 cM (LOD score = 2.1) explaining 4% of tube colour variation (data not shown), similar to the value explained by *WD40* genotype in linear regression in the V1640 population. Therefore *WD40* remains a candidate for a minor-effect QTL.

Ideally the role of candidate genes would be tested by complementation of *A. majus* mutants (preferably carrying null mutations) by comparing the effects of the different species alleles on anthocyanin pigmentation in *Antirrhinum*. This was done by transient expression of *Ros1* and *Ros2* in *ros* mutant petal cells by Schwinn *et al*. (2006). Such experiments were not possible in this project due to limited time.

Chapter 5

The attempts to find the novel colour gene.

In the two previous chapters, I tested the involvement of known ABP genes in differences in pigmentation between *Antirrhinum* species. The results suggested that a major effect QTL in LG7, which specifically affecting flower lobe colour, did not correspond to a known candidate genes. Therefore one way to identify the gene would be to fine map the QTL. In this chapter, several attempts were made to narrow down the location of the QTL relative to molecular marker in the linkage map. This included investigating the use of transposon insertions as molecular markers, with the possibility that the QTL mutation had been caused by transposon insertion.

5.1 Transposon Display (iPCR) approach

There are many ways to isolate genes from plants, but since the first isolation of transposons from Antirrhinum, they have proved useful for gene cloning. Insertion of a transposon into a gene can impair or change its structure, causing a mutation, but somatic and germinal excision can allow the mutation to revert to wild-type (Schwarz-Sommer *et al*., 2003; Luo *et al*., 1991; Lister *et al*., 1993). This behavior can be used to isolate mutated genes using the technique known as transposon tagging. Once the transposon insertion that is responsible for a mutation has been identified; the mutated gene into which the transposon has inserted can be isolated. There are a number of ways of identifying the transposon insertion. These include Southern hybridization with transposon probes and PCR based methods. One PCR-based method is a technique known as Transposon Display (TD) - a modification of the AFLP (Amplified Fragment Length Polymorphism) technique. This involves amplifying part of the flanking sequence of each transposon up to a specific restriction site, to which adaptors have been ligated. The resulting PCR products can be analysed in a high resolution polyacrylamide gels or an automated sequencer (Van den Broeck *et al*., 1998). The insertion can be detected by bands in the mutants that are absent from wild-type progenitors, that are

linked to the mutant phenotype in populations segregating for the mutation and that are lost on reversion of the mutation to wild-type (Van den Broeck *et al*., 1998). An alternative method of TD, developed by Souer and his colleagues (Souer *et al*., 1995), uses inverse Polymerase Chain Reaction (iPCR) with primers located only within the transposon sequence. This method has the advantage of allowing excision and purification of the candidate band prior to cloning and sequencing.

Inverse PCR (iPCR) was designed to amplifying anonymous flanking genomic DNA regions. The technique involves the digestion of source DNA, circulation of restriction fragments by self-ligation, and amplification using oligonucleotides that prime the DNA synthesis directed away from the core region of a known sequence, i.e. the opposite of the direction to primers used in normal PCR (Figure 54).

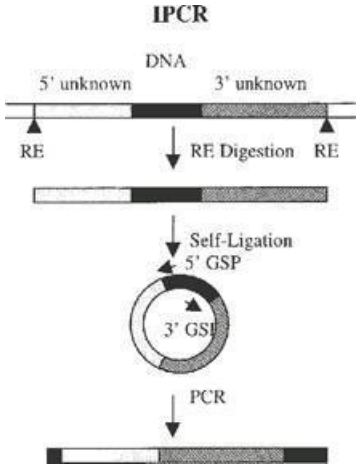


Figure 54. iPCR for cloning genomic DNA
The procedure consists of three steps: digestion of genomic DNA with a restriction enzyme (RE), circularization and amplification. The black and open bars represent the known and unknown sequence regions of double-stranded cDNA, respectively; GSP: gene-specific primer. (From Jong *et al*, 2002)

Many classical anthocyanin pigmentation mutants in *Antirrhinum* were result of transposon insertion into ABP genes such as *Niv*, *Pal*, *Del* (Bollmann, Carpenter and Coen, 1991; Coen, Carpenter and Martin, 1986; Goodrich *et al*., 1992). I had also found that *Ros1* alleles from different species differed by an insertion of *Tam661* in the 5'UTR. It is possible therefore that the LG7 QTL carries a transposon that can be detected by Transposon Display.

Observation of the F_2 families generated from *A. majus* x *A. molle* hybrids showed surprising variation that was difficult to explain by Mendelian segregation (David Greenshields, pers. comm.). A dark F_1 mutant (darker than other F1 hybrids; called 150) gave F_2 progeny with a large proportion of less pigmented plants (8:45:10; pale:intermediate:dark). In the meantime, a pale mutant F_1 (paler than other F_1 hybrids; called 109) gave F_2 progeny with a different proportion of pigmentation phenotypes (30:27:3; pale:intermediate:dark). One of the possibilities to explain this observation is the mutations are caused by transposons and are able to revert.

The *A. majus* parental line (JI.98) is known to carry the active transposable element Tam3, so it is plausible that the transposon is responsible for the mutations found in these families and for their reversion. To test this hypothesis, I used a transposon display (inverse PCR; iPCR) method with *Tam1*, *Tam2* and *Tam3* primers to try to identify the presence of new transposon insertions in the mutants and to test whether they co-segregated with pigment phenotypes. The method was used to compare F_2-progeny of the darkly pigmented mutant (A150) that had arisen in the F_1 of a cross between *A. majus* (dark flowers) and *A. molle* (pale flowers) and progeny of the pale flowered F_1 mutant (A109). The progeny of a normally pigmented F_2 plant (A118) were used as a control.

Pale and dark flowered progeny were selected from each parent and DNA extracted from a pool of each. I used *Mse*I and *Pst* I restriction enzymes to digest the genomic DNA. The method was expected to amplify multiple DNA fragments, each corresponding to a different transposon insertion. However, it proved to be very unreliable and different patterns of bands were seen in independent replicates of the same DNA sample. These were observed even when same PCR reagents and template DNA were used, suggesting that PCR was inconsistent (Figure 55).

There are several explanations for the unrepeatable results. One possibility is that the amount of template for each reaction was not enough to represent the whole genome.

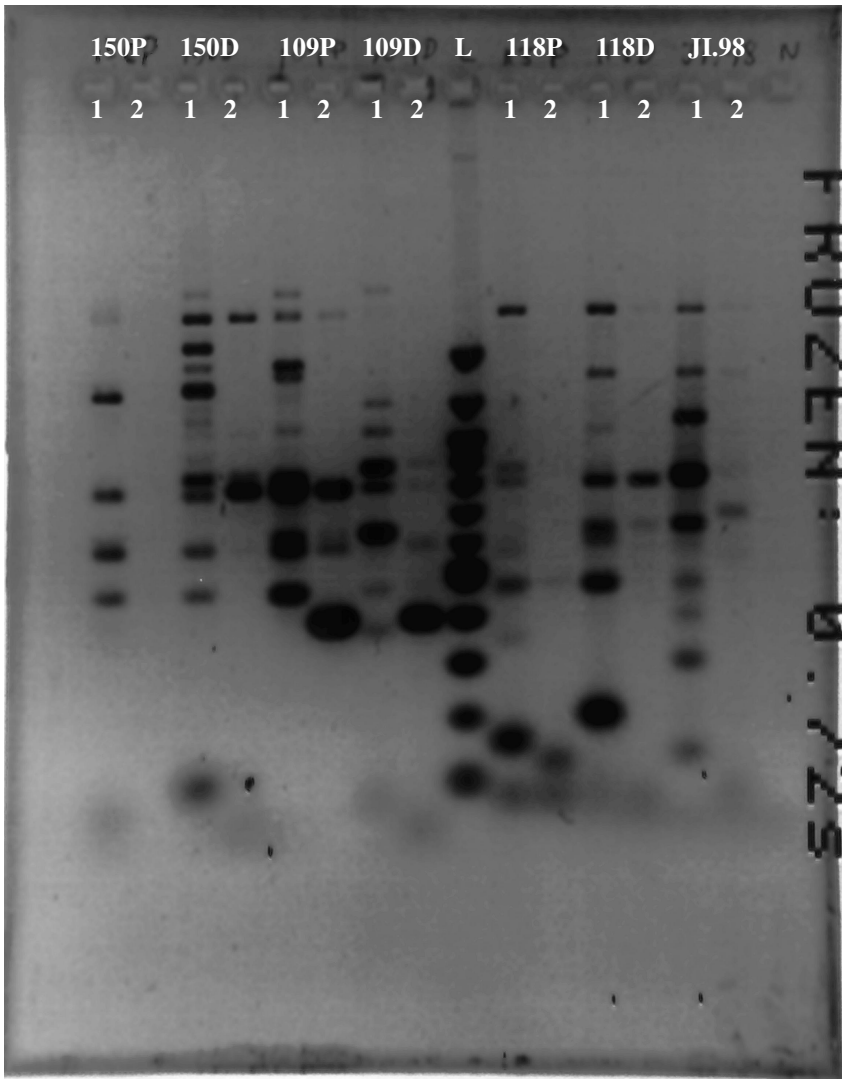


Figure 55. Transposon Display by iPCR using *Pst* I and Tam2 primers showing inconsistent results Numbers 1 and 2 indicate reactions that were done at the same time, using the same digested and ligated genomic DNA template. (150D = a pool of members of F2 family 150 with dark flowers; 150P = their siblings with pale flowers; 109D = dark members of family 109; 109P = their pale siblings; 118D & 118 P, dark and pale members of family 118, respectively; JI.98 = *A. majus* parent; N2 = PCR without genomic DNA template. L = DNA marker ladder)

However, when a larger volume of template was used, the results were still inconsistent (data not shown).

An alternative explanation was that the *Pst* I restriction enzyme had not completely digested the DNA. I therefore tried *Mse* I because this enzyme was used successfully to perform AFLP in Antirrhinum, suggesting that it can cut genomic successfully. Again the results proved to be inconsistent (data not shown).

I therefore chose other types of molecular markers to further map the LG7 QTL.

5.2 Mapping with amplified fragment length polymorphisms (AFLPs)

AFLP is a PCR-based technique developed in 1995 by Vos and his colleagues (Keygen N.V, Netherlands). It can amplify sequences around a particular restriction site and the number of amplified fragments can be controlled by selecting restriction enzymes with different number of sites in the genome and by using amplification primers with nucleotide extensions that are specific to only a subset of templates. It has the advantage that it can detect multiple polymorphisms without the need for sequence information. Because most of the AFLP fragments will correspond to unique positions on the genome, assuming that the total number of fragments is not too high, they can be used as markers in genetic and physical maps.

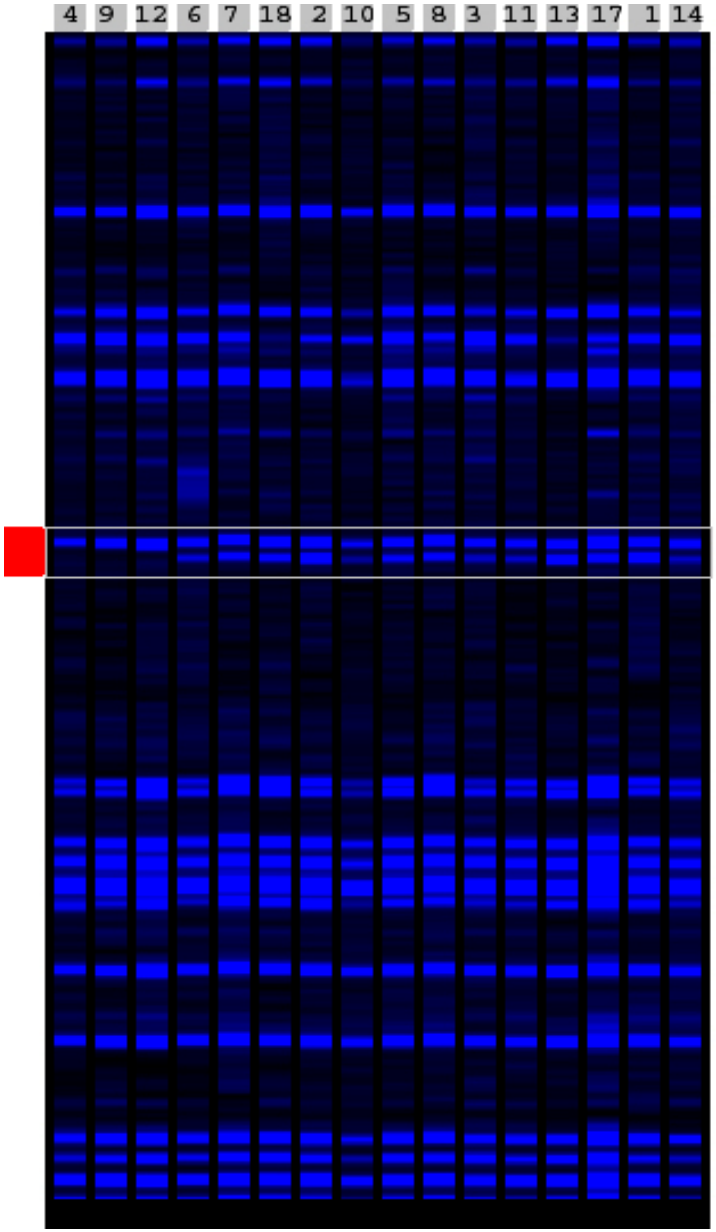
AFLP was chosen for the first attempt to map the LG7 QTL because relatively few gene based markers were available for LG7 and their relative positions in the chromosome were doubtful. Because the genetic background of the NIL M239, which is heterozygous at the QTL, was mainly from *A. majus*, it provided an ideal family to detect association between an AFLP and the QTL.

All individuals from NIL family M239 were tested by AFLP, together with two pale individuals from other NILs families (the palest from M241 and the palest from M244). DNA was digested with *Pst* I and *Mse* I and a number of different combinations of selective primers used to amplify different subsets of *Pst* I-*Mse* I fragments. The selective *Pst* I primer was labeled with a fluorescent dye to allow AFLP fragments to be separated in an ABI 3730 sequencer and virtual AFLP gels were constructed from the files generated by the sequencer. These gels showed similar patterns of bands in different individuals, as expected due to the similar genetic background of the NILs families. Some bands were found to segregate, but were not associated with flower colour in M239. Nevertheless, I detected one AFLP that was highly associated with flower colour in M239, as shown in Figure 56.

This AFLP is referred to as P11AGA256, because it was 256 bp in length and generated with selective *Pst* I primer P11 and an *Mse* I primer with the trinucleotide extension AGA. It is absent in all pale lines and present in all dark lines. Because the *A. charidemi* QTL allele is associated with pale flowers and is recessive with respect to the *A. majus* allele, pale flowered plants are expected to be homozygous for the *A. charidemi* allele. Therefore P11AGA256 is likely to have been generated from the *A. majus* chromosome. Dark flowered plants were therefore expected to be either homozygous or heterozygous for the A. majus allele and AFLP marker. Comparing P11AGA256 with the slightly larger band, which did not segregate, allowed its relative intensity to be estimated in the dark plants (Figure 58b). As suggested by van Eck and his colleagues (Eck *et al.*, 1995), the intensity of an AFLP band can indicate the copy number – i.e., a higher peak corresponds to plants homozygous for the *A. majus* alleles of P11AGA256 and a lower peak to heterozygotes. On this basis, family M239 consisted of *A. majus* homozygotes, heterozygotes and *A. charidemi* homoozygotes in a ratio of 3:10:3 that is not significantly different from the expected 1:2:1 ratio.

Because the linked AFLP was present in *A. majus*, which had been used as a parent in several mapping populations, previous maps were examined for an AFLP that had the same size and had been amplified with the same primers. An AFLP fitting these criteria was found in a map made from two *A. majus* parents (Schwarz-Sommer *et al.*, 2010), and located towards one end of LG7. This result confirmed that this AFLP marker was most likely linked to second major effect QTL.

The interpretation of band intensity as allelic copy number was further tested by sowing the selfed seeds of one of the plants that had been scored as heterozygous (plant M239.6). Its progeny segregated for flower colour, as expected, though the range of colour variation was not as large as in M239 (Figure 57), possibly because the plants were grown under different conditions.



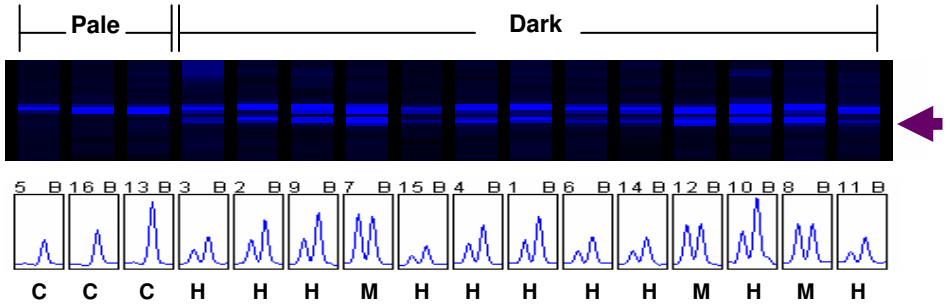


Figure 56. AFLP analysis of NILs family M239 (a) Part of the AFLP gel (100-300 bp for the primer combination P11-MAGA. Lane's numbers represent plant identities, with pale flowered plants in the first three lanes and dark plants from the fourth lane onwards. The highlighted region contains two bands, P11AGA256 (lower band) is segregating and P11AGA257 (higher band) is not. (b) The highlighted region with peak intensities for the two bands shown - P11AGA256 to the left and P11AGA257 to the right. P11AGA256 is absent (C) from all pale plants and either heterozygous(H) in dark plants (if P11AGA256 was present at about half the intensity of P11AGA257) or homozygous (M) (present at about the same intensity).

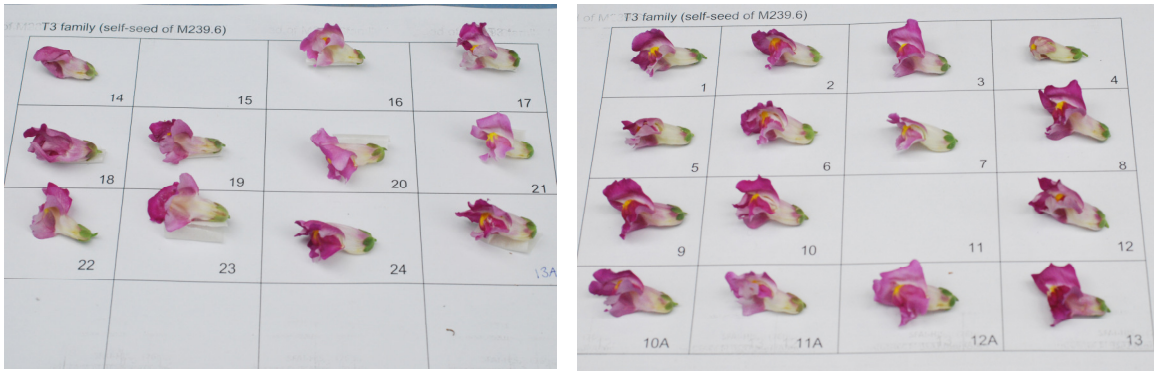


Figure 57. Flower colour variation found in the progeny of M239.6 The parent had been scored as heterozygous for P11AGA256.

The AFLP analysis also showed the palest individuals from two other NILs families (M241 and M244), which also segregated the LG7 QTL, also lacked P11AGA256, consistent with linkage of the AFLP to the *A. majus* allele of the QTL (data not shown). Therefore additional pale members of these two families were AFLP genotyped. The P11AGA256 band was found to be absent from 25 pale individuals, but present and heterozygous in another 5 of them. This can be explained by recombination between the AFLP and the QTL. On this basis, the AFLP marker was estimated to be about 8.3 cM from the QTL (5 recombinants in 60 chromosomes).

5.3 Mapping markers test in LG7

Having identified an AFLP marker linked to the LG7 QTL, additional markers were tested for linkage. A few gene based markers had previously been mapped to LG7 (Figure 16b, blue bold letters), although their relative positions were uncertain.

The gene based marker tested was *MIXTA-Like1* (*ML-1*), which had been mapped at 39 cM (Figure 16b, Chapter 3). *MIXTA* encodes a MYB-related transcription factor that is expressed in adaxial flower petals necessary for the development of conical cells on the adaxial petal epidemis (Glover *et al*., 1998). It is believed that conical cells increase the

proportion of incident light that goes into to epidermal cells, enhancing light absorption, hence, increase the apparent intensity of flower colour (Noda *et al*., 1994; Figure 58). A similar effect might account for the role of the LG7 QTL in flower colour. However, *ML-1* was reported to be expressed strictly in ventral petal of the corolla (Perez-Rodriguez *et al*., 2005), so was unlikely to be a candidate for the QTL.

Previous PCR and sequencing in *A. majus* and *A. molle* had shown that there are two very similar *ML-1* genes, *ML-1a* and *ML-1b* that are closely linked to each other (Suxin Yang, pers. comm.). The first and second exons, which encode the conserved MYB domain, were identical in both paralogues. The *ML-1b* had been mapped to LG7 as a co-dominant CAPS marker in the *A. majus* x *A. charidemi* F_2 population, V1604 (Feng *et al*., 2009).

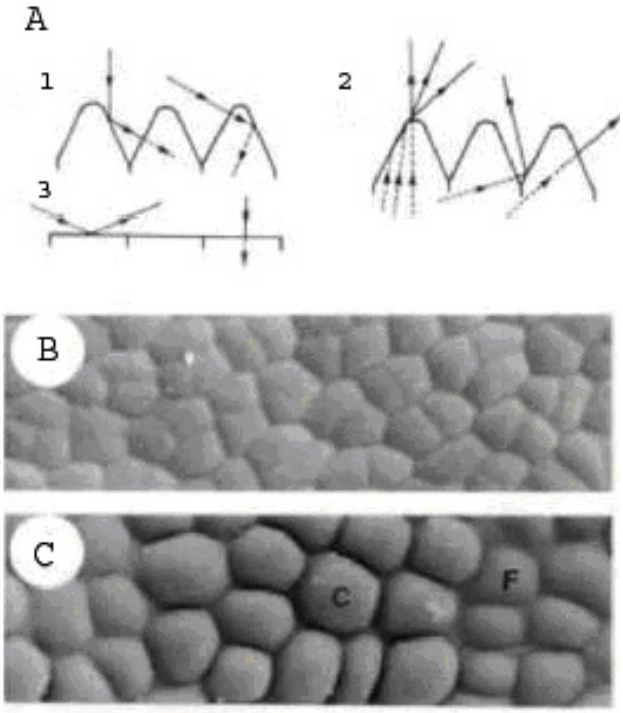


Figure 58. Conical shaped epidermal cell affect flower pigmentation (pictures and description from Noda *et al*., 1994). (A) Incident light is reflected or refracted into cells by their angled cell walls (1). Light reflected out from the mesophyll will be refracted in many directions (2), giving the epidermis a bright appearance from many angles. This can enhance the intensity of flower colour, compared to flat cells, which mainly reflect light before it has entered the cell (3). Epidermal cells under the SEM (Scanning Electron Microscope) of a *MIXTA* mutant (B) compared to wild-type (C).

The same co-dominant CAPS marker for *ML-1*, involving the presence or absence of a *Hpa* II site, was used to genotype members of family M239 and found to show complete co-segregation with flower colour (data not shown). Plants homozygous for the *A. charidemi ML-1* allele (*AcML-1*) were all found to be pale, whereas the dark siblings were homozygous for the *A. majus ML-1* allele (*AmML-1*) or heterozygotes. However, one recombinant (a pale heterozygote) was found in NIL family M241 among 14 pale plants, though all 10 pale plants in M244 were homozygous *AcML-1*. The recombination rate between the QTL and marker was estimated to be around 1.9 cM (one recombinant chromosome in 54).

Although *Inc* had been eliminated as a candidate for the QTL, it had been found to show linkage to it. Genotyping NIL family M244, which segregated for both LG3 and the region of LG7 including both the QTL and *Inc*, revealed that two out of 11 pale individuals were heterozygotes. Thus, the estimated distance of *Inc* from the QTL was about 9 cM (two recombinant chromosomes in 22).

The *AmAnt* marker had been mapped to 42 cM in the *A. majus* x *A. charidemi* F_2 population V1064 by Feng *et al*., 2009, as a dominant *Rsa* I CAPS polymorphism that identified the *A. majus* allele. The same polymorphism was used to genotype the NILs familiy M244. Six out of 11 pale plants were found to carry at least one *A. majus AmAnt* allele, suggesting that *AmAnt* is not closely linked to the flower colour QTL. Further evidence for this was provided by NILs family M241, which did not appear to segregate for the LG7 QTL, or the other markers that had been found to be closely linked to it, but did segregate for *AmANT*. In both families, fewer plants were found to carry the *A. majus* allele than expected (~50%, compared to 75% expected). This can be explained by linkage of the *A. majus* allele to a recessive mutation that is zygotic lethal in this background, so that *A. majus* homozygotes are not recovered. If this is the case, the observation that flower colour segregates in the expected 3 dark: 1 pale plant in family M241 further supports the idea that *AmANT* is not closely linked to the QTL.

The marker *FAP2* was tested next, its alleles distinguished by a *Hae* III CAPS polymorphism. Thirteen out of the 14 pale members of M241 and nine of 11 pale members of M244 were homozygous for the *A. charidemi FAP2* allele. This gives an estimate of 6 cM between *FAP2* and the QTL.

The last marker tested was *5.08.o20*. Although a CAPS polymorphism segregated in both M241 and M244, only four pale plants were found to be homozygous for the A. charidemi allele, and therefore this marker behaves as if it is unlinked to the flower colour QTL.

The genotypes of pale plants with recombinations between at least one of the molecular markers and the QTL were used to estimate a map of this region of LG7 (Figure 59). This will be useful in further fine mapping of the QTL.

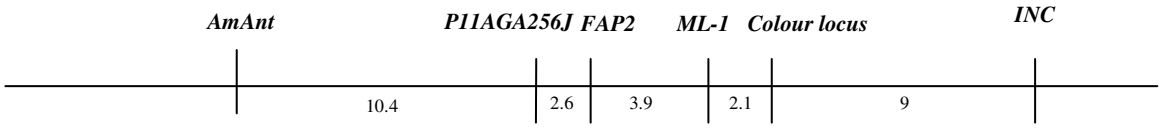


Figure 59. A linkage map of the region around the LG7 flower colour QTL estimated from NILs families. Estimated map distances are shown as recombination percentages, without correction.

5.4 ABP gene expression

Although *Ros* had been implicated as a QTL underlying variation in corolla tube colour between *A. majus* and *A. charidemi*, no candidate had been found for the QTL in LG7, which had a major effect on lobe pigmentation. To investigate whether this QTL was likely to be a structural gene or a regulatory gene, RNA was extracted from pale and dark members of NIL family M239, which segregate for this QTL, but not at *Ros* or *Inc*.

If the LG7 QTL is regulatory, pale flowers were expected to show lower levels of late-pathway structural gene expression compared to dark flowers. If the QTL is a structural gene, then it is less likely to affect expression of other ABP genes.

Two fully developed flowers were taken from each plant, photographed and their corolla lobes frozen in liquid nitrogen. I did not test expression in corolla tube because the LG7 QTL affected lobes predominantly and family M239 had generally paler tubes, consistent with its homozygosity for the *A. charidemi* allele at *Ros*. Expression was compared by RT-PCR in two dark flowered plants (M.239.1 and M239.2) that were likely homozygous for the *A. majus* allele of the LG7 QTL, based on genotypes of flanking markers, one putative heterozygote (M239.6) and one pale flowered plant that was probably homozygous for the *A. charidemi* allele (M239.4) (Fig. 60c). Expression of the *AMA2* gene, which acts as a control for RT-PCR, showed saturated amplification in the pale NIL making it difficult to compare expression of other anthocyanin-related genes in a quantitative manner. However, some of these genes showed obvious difference between pale and dark flower, i.e. *Pal*, *PAT1*, *Mut* and *Ros2*. Expression was clearly higher in pale flowers for the structural gene *Pal*, the putative anthocyanin transporter gene *PAT1* and the regulatory *Mut* and *Ros2* genes. However, the expression level of the *Niv* and *Inc* genes showed inconsistency in two different dark flowers and both *Niv and Inc* possibly showed higher expression in pale flowers (homozygous for the *A. charidemi* QTL allele) compared to dark flowers (homozygous for the *A. majus* QTL allele). The possibility of higher level of expression in pale flowers also included *Eos*, the regulatory *WD40* and *Del*. Expression of *PAT2* and *PAT4* was not detected in any of the samples (data not shown), suggesting that these genes are not expressed in opening flowers.

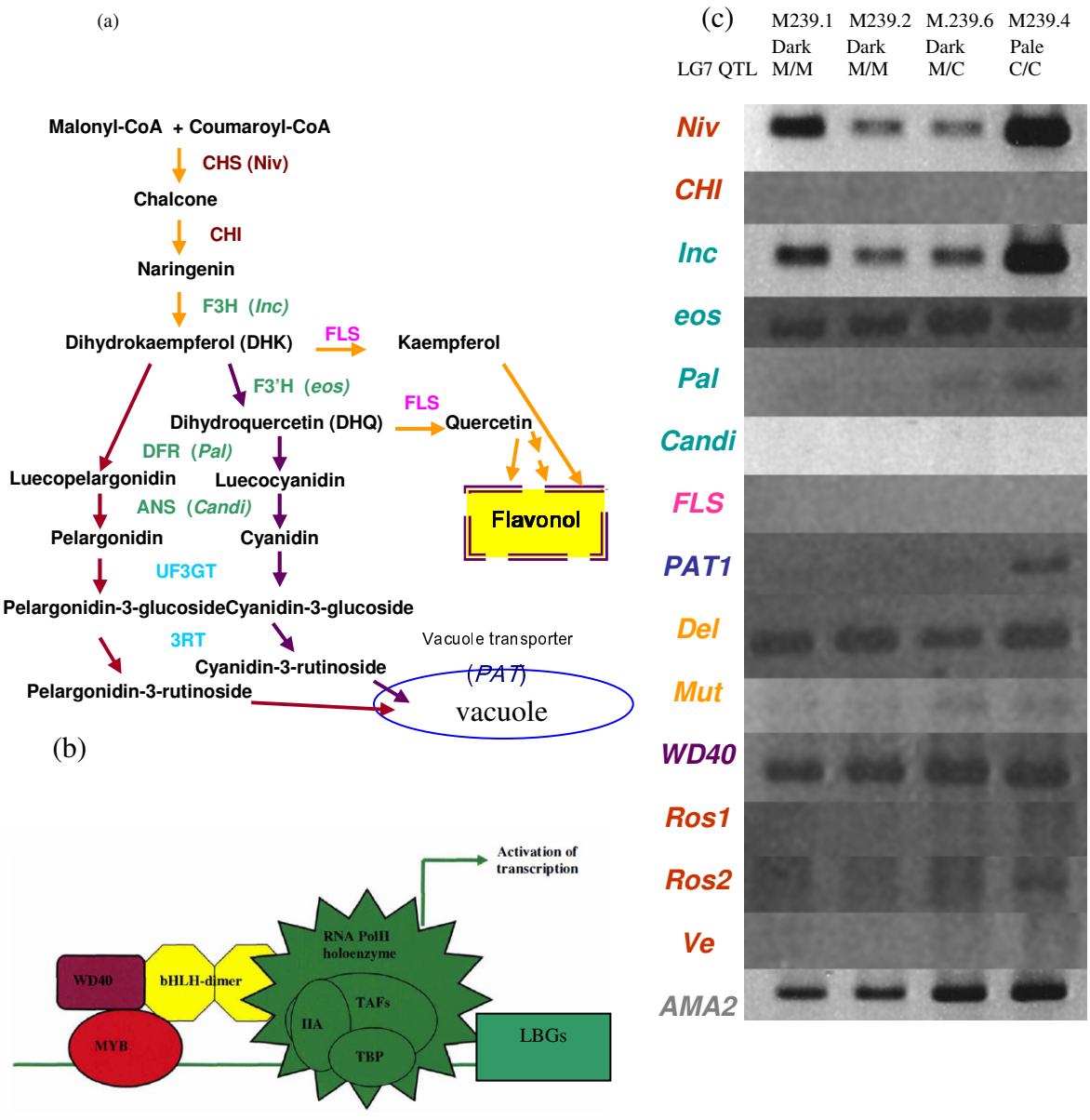


Figure 60. Expression of genes involve in anthocyanin biosynthesis. (a) A schematic diagram showing anthocyanin production in *Antirrhinum*. (b) The bHLH-WD-MYB regulatory complex that can affect expression of late biosynthetic genes (LGBs). (c) RT-PCR of genes involved in anthocyanin synthesis. Early biosynthetic genes (EBGs; *Niv* and *CHI*) do not respond to regulators of the anthocyanin biosynthesis pathway are shown in brown, late genes, which do respond are represented in green. *Fls*, which might divert substrate from the ABP, is in pink. Putative anthocyanin transporter genes (*PAT*) are represented in blue. The bHLH regulatory genes are in yellow, the *WD40* gene in purple and *Myb* genes in red. *AmA2* represent a constitutively expressed control (Grey). Lane 1 = M239.1 (dark flower); Lane 2 = M239.2 (dark flower); Lane 3 = M.239.6 (putative heterozygous of LG7 QTL); Lane 4 = M239.4 (pale flower).

Moreover, some structural and regulatory gene expression, i.e. *CHI, Candi*, *FLS* and *Ros1*, were very low or undetectable in this experiment. This was not unexpected for *CHI* and *Candi*, which are expressed only at a very low level in expanding flowers (Jackson *et al*., 1992). Low expression of *Ros1* was consistent with homozygosity of the *A. charidemi Ros1* allele, which is expected to be expressed at a lower level than the *A. majus* allele. There was also no detectable *Ve* expression as expected from the lack of a darker venation phenotype in these families.

Differences in the primer binding sequences between alleles can be ruled out as an explanation for the differences in amplification in the different genotypes in RT-PCR, because the primers were designed to match both the *A. majus* and *A. charidemi* alleles perfectly and were able to amplify both alleles with similar efficiency from genomic DNA (data not shown). The results were therefore unexpected in three ways: 1) they suggested that lower anthocyanin levels correlate with higher expression of genes needed for anthocyanin biosynthesis; 2) that expression of both structural and regulatory genes were affected and 3) that different genes were affected in different ways (e.g., *Inc* showed little variation, whereas *Pal* expression was significantly higher in pale flowers).

One explanation for the effect of the LG7 QTL is that it changes the flow of precursors into the anthocyanin and flavonol pathways. To test whether competition from flavone synthesis might be responsible for differences in anthocyanin levels, expression of *Flavonol synthase* (*FLS*) was examined. No expression could be detected in any of the flowers (Figure 60).

Comparing the expression in plants homozygous for one of the LG7 QTL alleles with heterozygotes also suggested different dominance effects for the QTL. The *A. majus* allele has a dominant effect on anthocyanin biosynthesis because *A. majus* homozygotes are indistinguishable from heterozygotes – i.e., the pale allele is fully recessive, as expected of a loss-of-function mutation. Similarly the *A. majus* QTL allele appeared to have a dominant effect in repressing *PAT1* and *Ros2* expression, which were detected at

high levels only in *A. charidemi* homozygotes. For *Pal* expression, and possibly other genes showing intermediate expression in heterozygotes, the QTL alleles appeared to be semi-dominant, while for *Mut* and *Del* expression, which was similar in the *A. charidemi* homozygote and heterozygote, but lower in the *A. majus* homozygotes, the *A. majus* allele appeared to act as a recessive suppressor of expression.

To investigate the effects of the QTL on ABP gene expression further, a pale plant was selected from family M241 (M241.148) and petals sampled from young flowers buds, between 15-20 mm long, and from mature, open flowers,. Mature petals were also harvested from a pale member of M244 (M244.160) and a pale flowered progeny of M239.6 (T3.21) and mature petals from two of its dark flowered sibs. RT-PCR was repeated under the same conditions as before, except that for the genes *CHI*, *FLS*, *Candi*, *Ros1* and *Ve*, which showed low or undetectable expression previously, the number of amplification cycles was increased from 30 to 35. Results are shown in Figure 61.

In different NILs families, *CHI*, *FLS*, *Eos*, *Candi*, *PAT1*, *Mut*, *WD40*, *Ros1* and *Ros2* showed inconsistent levels of expression between pale lines and dark lines. However, some genes still showed an inverse correlation between expression and colour, i.e. *Niv*, *Inc*, *Pal* and *Del*. With a higher number of amplification cycles, *Ve* was found to have saturated levels of expression in both young and mature flower, whether dark or pale, suggesting that the number of amplification cycles is too high to compare expression levels between flowers and therefore the association of this gene to flower colour variation in the NILs remains unknown.

Ros2 expression was also inconsistent. Expression was detected in two plants homozygous for the *AcRos* allele (M244.160 and T3.21). However, two dark plants which were also *AcRos2* homozygotes (T3.25 and T3.28) had no detectable expression.

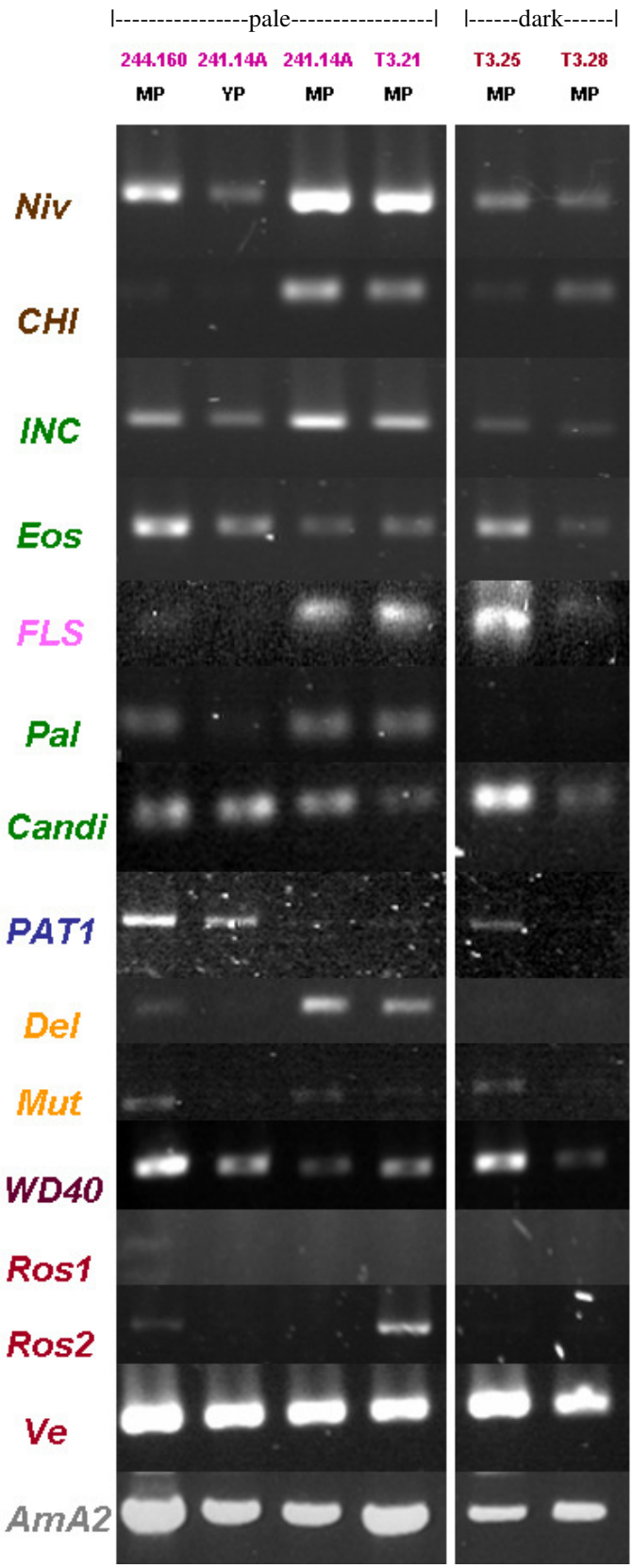


Figure 61. Expression of ABP genes in families segregating for flower colour QTL Early biosynthetic genes (*Niv* and *CHI*) that do not respond to known regulators of the anthocyanin biosynthesis pathway are shown in brown, late genes, which do respond are represented in green. *Fls*, which might divert substrate from the ABP, is in pink. The putative transporter PAT1 is in blue. The bHLH regulatory genes are in yellow, WD regulatory gene in purple and *Myb* genes in red. *AmA2* represents a constitutively expressed control (Grey). Lane 1-4 are pale flowers at different developmental stages and from different families. Lanes 5 & 6 are dark flowers from the same family. MP = mature petal, YP = young petals. Dark and Pale expression analysis of each gene was done in the same gel.

Amplification of the actin control gene, *AmA2*, also appear to be saturated in one of the pale NIL plants (M244.160), which make it difficult to compare expression of other genes between flowers of this plant and others. The results could be evaluated better by repeating the experiment under conditions in which expression of *AmA2* was consistent between lines. However, this was not possible in the project time frame.

Expression between unopened, young flower buds (15-20 mm in length) and mature opening flower (> 30 mm in length) differed. According to Jackson and his colleagues (1992), expression of most structural genes (*Niv*, *CHI*, *Inc*, *Pal*, *Candi* and *UFGT*) peaks in *A. majus* flower buds when they are about 15-20 mm long. In this experiment, however, it was obvious that *Niv*, *CHI*, *Inc*, *Pal, FLS* and *Del* were expressed at higher levels in mature flowers whereas *Eos, PAT1* and *WD40* were expressed at lower levels (Figure 61, Lane 2-3). It was therefore possible that the second major-effect QTL affected the timing of anthocyanin gene expression, so that the *A. charidemi* allele delayed maturation and hence anthocyanin production, resulting in paler flowers.

5.5 Testing whether QTL affect flavonol:anthocyanin ratios

As discussed previously, one explanation for the effects of the LG7 QTL is that it is involved in an unknown regulatory system that promotes flavonol biosynthesis at the expense of anthocyanins and that more flavonols are produced in paler flowers homozygous for the *A. charidemi* QTL allele. Increased ABP gene expression might then result from loss of negative feedback from the anthocyanin pathway (i.e. anthocyanin products repressing ABP gene expression so that reduced levels of products enhances expression) or positive feedback from flavonol synthesis (i.e. accumulation of flavonols enhancing ABP expression). To test whether the QTL affected the relative proportion of flavenols and anthocyanins was possible, I compared the absorption spectra (190-600 nm) of pigments extracted from pale and dark corolla lobe tissue in acidic methanol (MeOH). The prediction was that if the LG7 QTL affected flavonols and anthocyanin synthesis in opposite ways, pale lines would have a higher

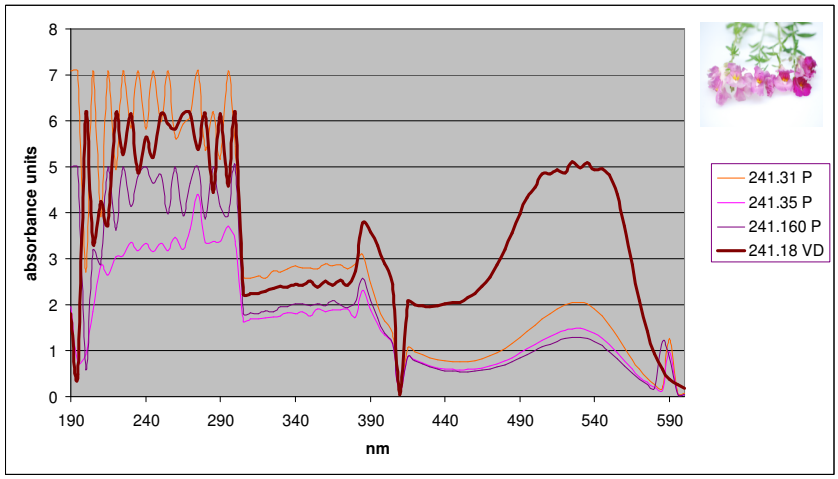
flavonol:anthocyanin ratio than dark plants. The major flavonols were expected to be kaempferol (with maximum absorption at 266 nm, 322 nm and 366 nm in acidic MeOH), and quercetin (absorbing maximally at 256 nm 374 nm; Guttuso *et al*., 2007). However, flavonol and anthocyanin absorbance peaks can shift slightly depending on pH and metal ions. Therefore the ratio of kaempferol or quercetin was estimated from absorbance at 255 nm 265 nm, respectively, and compared to total anthocyanin at 530 nm (Kim *et al*., 2008), rather than attempting to quantify flavonols and anthocyanins separately.

In M241, in which flower colour variation was related to segregation of LG3 and not LG7, the ratio of kaempferol:anthocyanin (kae:ant) and quercetin:anthocyanin (que:ant) was between 2.1-3.4 (kae:ant) and 2.1-3.0 (que:ant) in pale flowers, whereas in dark flowers it was around 1.2 for both kae:ant and que:ant (Figure 62a). Thererfore the LG3 QTL appears to affect flavonol and anthocyanin levels disproportionately.

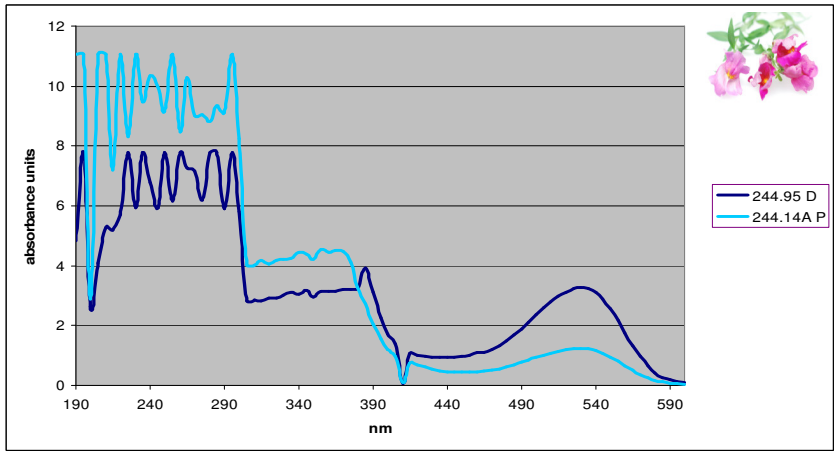
Pale flowered members of family M244, segregating for LG7, had much higher kae:ant and que:ant ratios (kae:ant 8.8, que:ant 8.2) than their dark flowered siblings (kae:ant 1.9, que:ant 2.2; Figure 62b). The same trend was also found in the progeny of a member of M244 that segregated for the QTL (Figure 62c).

Results from this preliminary test were consistent with a role of the LG7 QTL in regulating flavonol biosynthesis at the expense of anthocyanin production. Unfortunately, the time did not permit this to be investigated further. The method has a major disadvantage in that it cannot determine whether the absolute amount of flavonoids is altered by the QTL or whether the different ratios reflect only lower anthocyanin production. Therefore the findings do not rule out other explanations for the role of the QTL.

a)



b)



c)

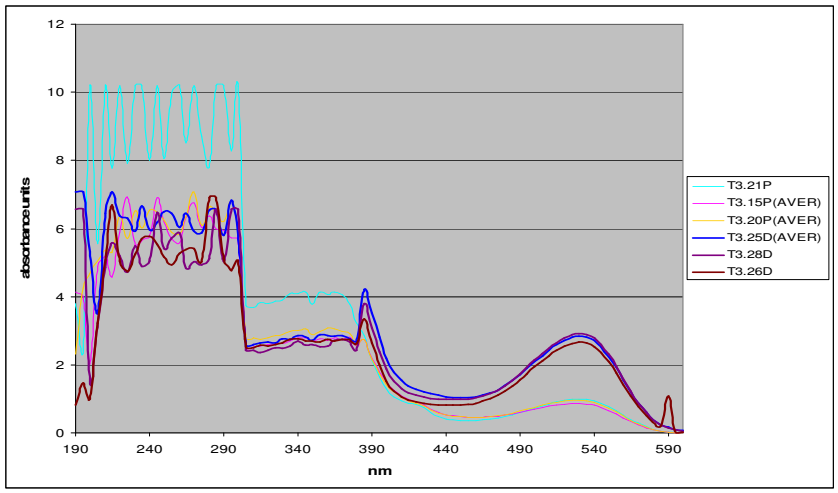


Figure 62. Absorption spectra of petal extracts from plant segregating flower colour QTL (a) In family M241 (segregating LG3) the ratio of flv:ant was lower than in other NILs families. Flowers of the plants tested are shown next to the spectra, 131 (pale), 135 (pale), 160 (pale) and 18 (dark) from left to right. (b) Family M244, segregating LG7; 14A (pale; left) 95 (dark, right). (c) The progeny of a heterozygous member of family 244.

5.6 Conclusion and Discussion

Transposon display was used in an attempt to find transposons linked to the LG7 QTL. However, this proved unreliable and was not pursued further. AFLP analysis was then used in NILs and identified a marker, P11AGA256, about 9 cM away from the QTL. Testing other markers known to map to LG7, I was able to map the QTL to an interval of around 11 cM, between the markers *ML1* and *Inc*. In the future these markers can be used to identify recombinations around the QTL, which could be used to map it. Currently the *Antirrhinum* genome sequence is not fully assembled and consists of several thousand sequence scaffolds. *ML1* and *Inc* are present in different scaffolds, and therefore it is not possible to identify additional molecular markers or candidate genes in the interval from the sequence. However, this will be possible once the genome sequence is assembled fully.

One of the obvious questions about this novel colour gene is how it affects anthocyanin pigmentation - whether it is a structural gene, for example encoding an unknown anthocyanin-modification or transporter, or whether it is a regulatory gene that affects expression of ABP structural genes. To address this question, its effects on expression of known ABP genes were examined.

The results of ABP gene expression analysis in members of NILs families segregating for the LG7 major-effect QTL were unexpected. Pale flowers appear to express the regulatory genes *Ros2* and *Mut* at higher levels than dark flowers. Although this is consistent with the idea that *Ros2* might act as a suppressor of anthocyanin synthesis, *Mut* is known only to promote anthocyanin gene expression and *Ros2* can activate anthocyanin biosynthesis in transfected petal cells. Similarly, the structural genes *Pal* and *PAT1* were expressed at higher levels in pale flowers. Independent RT-PCR with other NILs plants with pale flowers suggested that *Niv*, *CHI*, *Inc* and *Eos* were also expressed at higher levels than in their dark flowered siblings (Figure 60c and Figure 61). These results suggested that the *A. charidemi* allele of the QTL increases

expression of structural genes throughout the ABP. *Ros1* and *Ros2* are known to be required for regulation of the late ABP gene, especially *Pal* and *Eos*, but not for the early biosynthetic genes such as *Niv* and *Chi* (Figure 63; Schwinn *et al*., 2006). Therefore the QTL is unlikely to act on structural ABP genes only by increasing *Ros* expression.

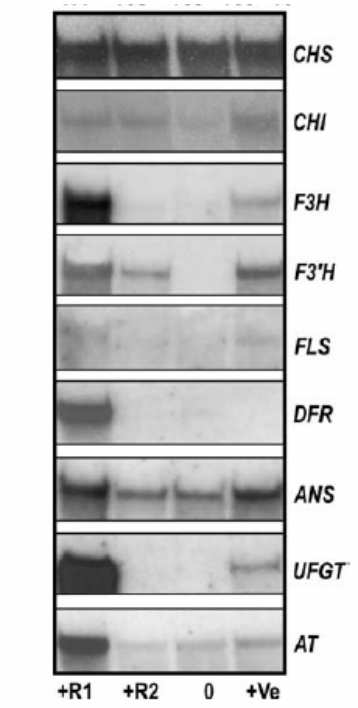


Figure 63. Northern blot analysis showing that mutations in the MYB-related regulatory genes, *Ros1*, *Ros2* and *Ve*, affect structural anthocyanin biosynthesis gene expression. +R1 is a wild-type *A. majus* line which carries active *Ros1* and an inactive *ros2* allele. +R2 is Ros^{dor} which carries active *Ros2* but inactive *Ros1*. 0 is the Ros^{col} line with reduced activity of *Ros1* and *Ros2*. +Ve is Ros^{dor} with active *Ve* (from Schwinn *et al*., 2006)

The increased expression of ABP structural genes in the pale NIL genotypes is similar to the up-regulation causes by a group of MYB-related proteins in Arabidopsis - the R2R3-MYB subgroup 7, which so far contains *MYB11*, *MYB12* and *MYB111*. This group of genes, referred as *PFG1-3*, was reported to control early flavonoid biosynthesis pathway by targeting genes encoding Chalcone synthase (CHS) , Chalcone isomerase (CHI), Flavonoid-3-hydroxylase (F3H) and Flavonol synthase1 (FLS1). The *pfg1-3* triple *myb* mutant showed flavonol accumulation in seedlings and slightly higher accumulation of anthocyanins than in the wild type, indicating that precursors usually channelled to flavonols are used to form anthocyanins (Stracke *et al*., 2007). Stracke and his colleagues also reported that another specific target of this MYB-subgroup 7 is UDP-

glucosyltransferase (*UGT91A1* and *UGT84A1*) required for producing kaempferol glucosides and quercetin glucosides.

It is therefore possible that the LG7 major-effect QTL encodes one of the R2R3 MYB subgroup 7 proteins, and that it might interact with *Mut* or *Del* in corolla lobe. In this case, the up-regulation found in the pale plants could be due to increase flow of precursors into production of flavonols (which are normally yellow in colour, but less intense than flavones), whereas the up-regulation of *Pal* and *Eos* occurred in response to increased *Ros2* expression to compensate for reduced anthocyanin production. The alleles of the QTL might have an additive effect on gene expression, explaining why the darker NILs plants that are homozygous for the *A. majus* allele (which in this case would have lower activity than the *A. charidemi* allele) show less expression of some ABP genes, while the heterozygotes have slightly higher expression than the *A. majus* homozygotes but slightly lower than the *A. charidemi* homozygotes.

FLS, which encodes the enzyme responsible for diverting precursors into flavonol synthesis, gave inconsistent estimates of expression in both pale and dark flowers, arguing against the anothocyanin-flavonol competition hypothesis. However, the R2R3 subgroup 7 proteins only up-regulate one of the *FLS* genes in *Arabidopsis* (*FLS1*; Stracke *et al*., 2007), raising the possibility that another regulated *FLS* gene is important in *Antirrhinum*. Although only one *Antirrhinum FLS* gene has been published so far (Schwinn *et al*., 2006; EMBL entry: DQ272591.1); searching the *A. majus* genome sequence, using Gapped BLAST (Altschul *et al*., 2007), reveals that it contains a small *FLS* gene family. One of the additional genes is more similar to Arabidopsis *FLS1* (Preuβ *et al*., 2009) than is the original *Antirrhinum FLS*. Therefore the possibility that *FLS* is the QTL or is regulated by the QTL remains.

Expression analysis in pale plants from NILs showed inconsistent levels of some genes, including *CHI*, *Eos*, *Candi*, *PAT1*, *Mut*, *WD40*, *Ros1* and *Ros2*. The inconsistent expression of *Candi* and *WD40* could not be attributed to variation in the genes

themselves, because neither *Candi* nor *WD40* genotypes showed any correlation with flower lobe colour variation and map outside regions that are segregating in the NILs. Because no polymorphism had been detected in *Eos*, it had not been mapped and therefore variation at *Eos* could not be ruled out as the cause of variation in its expression However, expression of *CHI* was interesting because it was undetectable in one of the pale plants (M244.160). There were two explanation for this; 1) this pale plant carried a weaker *CHI* allele, possibly the *A. charidemi* allele, and therefore produced less anthocyanin, or 2) this pale plant carried the *A. majus CHI* allele, which did not respond to a transcription factor (e.g., the proposed regulator of anthocyanin and flavonols), so that less flavonols and anthocyanins were produced. This idea could be tested further by genotyping the NILs plants at *CHI.*

The *PAT1* gene was proposed to encode a putative anthocyanin transporter in this project. It was included in the experiments because a similar gene (*PAT1*) in tomato was up-regulated when *Ros* and *Del* regulators were expressed in transgenic tomato fruits (Butelli *et al*., 2008). *PAT1* shows sequence homology to *TT12* of Arabidopsis, a member of the vacuole-localized multidrug and toxic compound extrusion (MATE) group (Mathews *et al*., 2003). Anthocyanin transport using MATE transporter is proposed to involve two processes. The first involves glutathione-S-transferases such as the products of the *TT19* gene in Arabidopsis (Sun, Li and Huang, 2011), *bronze-2* in maize (Marrs *et al*., 1995) and *an9* in petunia (Mueller *et al*., 2000). The second involves the MATE transporters such as encoded by *TT12* in Arabidopsis (Marinova *et al.,* 2007) and *MATE1* in *Medicago truncatula* (Zhao and Dixon, 2009). Involvement of a MATE transporter in flower pigmentation was recently confirmed in *M. truncatula* by knock-outs of *MATE2* causing pale flower (Zhao *et al*., 2011). Though I found no associated between *PAT1* and flower lobe colour variation in this project, the possibility remains that the up-regulation of *PAT1* was independent from flower colour variation in the lines I tested as the involvement of these anthocyanin transport systems might differ between tissues and species.

The transcriptional regulator *Mut* also showed inconsistent expression; one of the dark flowered plants showed higher expression, though no detectable expression of *Ros1* and another showed no detectable *Mut* expression. Since *Mut* is proposed to interact with *Ros1* in flowers, this suggests the possibility that *Mut* is required in other regulatory system.

Ve expression was detected at a saturated level in all plants with a higher number of PCR cycles. This was unexpected because none of the plants showed visible venation in lobe tissues. One explanation is that because the *Ve* region amplified in this expression analysis has no intron, that the products were amplified from genomic DNA contamination. If it was not the case, there are two possible alternatives; 1) that VE was expressed in veins but that its effect on anthocyanin synthesis was not apparent or 2) that *Ve* was generally expressed at a very low level but that the VE protein was not produced.

To investigate further whether flavonol production is up-regulated in pale flowers, preliminary tests of flavonoid and anthocyanin contents were conducted. The results suggested that flavonoid:anthocyanin ratios are higher in pale flowered plants. Though this implies that the QTL affects flavonoid and anthocyanin production in different ways, it does not reveal whether flavonoid synthesis is increased in pale flowered plants. This would require more accurate measurement of flavonoid content.

Chapter 6

Conclusion

In order to find the genetic differences that are responsible for flower colour variation between *A. majus* and *A. charidemi*, QTL analysis was carried out on an F_2 population. This suggest that two major-effect QTL affect flower colour variation in both corolla lobe and tube, located around 59 cM in LG3 and at 45 cM in LG7. Significant smaller effect QTL affecting flower lobe or tube colour variation were also detected, suggesting that different systems operate in different parts of the flower.

Although the *A. charidemi* parent of the mapping population was produced from seeds collected in the wild, the *A. majus* parent had been maintained for many generations of inbreeding as a laboratory line and had probably been derived originally from a horticultural cultivar. "Wild-type" cultivars of *A. majus* usually have more darkly pigmented flowers than their closest wild relative (*A. majus* subspecies *pseudomajus*). Therefore some of the identified QTL might represent mutations that were selected during the domestication of *A. majus*.

Previous studies (Scwhinn *et al*., 2006; Whibley *et al*., 2004; Stubbe, 1966), had implicated *Ros* in flower colour differences between *Antirrhinum* species. For *A. majus* and *A. charidemi*, *Ros* was linked to a QTL underlying variation in corolla tube colour and this linkage was confirmed in near isogenic lines segregating for part of LG3 around *Ros*.

In the *A. majus* x *A. charidemi* F_2 population, *Ros* was mapped at 64 cM in LG3 and this locus could explain about 26% ($R^2 = 0.2623$, P-value = 1.86×10^{-10}) of the variation in flower lobe colour and 50% ($R^2 = 0.5044$, P-value = 3.6×10^{-22}) of the variation in flower tubes. However another major-effect QTL for flower colour had been mapped to 59-60 cM in LG3 (Table 8, 9) and therefore the effects of these two QTL might have not been fully separated in QTL analysis in the F_2 population. Use of NILs families segregating loci around 64 cM confirmed that *Ros* was not the fully responsible for major-effect QTL

in LG3. The possibility remains that this QTL corresponds to the *Pal* gene, or possibly to the *Eluta* locus, which is known to be linked to *Ros* (Whibley *et al*., 2006).

Another explanation why *Ros* might have less of an effect than estimated for the LG3 QTL when segregating in NILs M241 and M244 is that the effect of *Ros1* is dependent on another gene and that this gene could have been lost by segregation during generation of the NILs. If this is the case, the interacting gene might encode a WRKY transcription factor, which are known to regulate *Ros* orthologues in Arabidopsis. From allelic variation analysis, I found a WRKY recognition site in *AmRos1* allele but not in *A. charidemi* and *A. molle* alleles. WRKY transcription factors were reported to have the association with *PAP1* expression and anthocyanin accumulation in Arabidopsis by Tohge *et al*. (2005) though no direct evidence about its regulation.

Ros homologues are proposed to have contributed to flower colour variation in other species. In *Petunia hybrida* inactivation of *AN2* causes white flowers (Quattrocchio *et al*., 1999) and several loss-of-function mutations in *AN2* were involved in the independent evolution of white flowers in other *Petunia* species, associated with a switch to pollination by nocturnal hawk moths (Hoballah *et al*., 2007). Related genes are also implicated in intra-specific variation in flower colour, flower colour patterns and leaf pigmentation in *Mimulus guttatus* (Lowry *et al*., 2011).

In *Antirrhinum*, *Ros* contains two tandemly repeated *Myb* genes. Variations in the structure of *Ros1* supported a role in flower colour differences between species. In *A. molle*, *Ros1* carried a frame-shift mutation that might lead to a loss of activity. This was consistent with previous allelism tests that had suggested that reduced *Ros1* activity contributed to the pale flowers of *A. molle* (Schwinn *et al*., 2006; Stubbe, 1966). A transposon was identified in the 5' UTR of *Ros1* in *A. majus* (lines JI7 and JI98) and *A. charidemi*. Type II transposons typically cause a short duplication of their host sequence on insertion and very rarely restore the ancestral sequence on excision, usually leaving a sequence "footprint" (reviewed by Colot, Haedens and Rossignol, 1998). The insertion in *A. charidemi* was 508 bp longer than in *A. majus*, as alignment showed that the internal

region of Tam661 was missing, suggesting either than a second insertion had occurred between *A. majus* and *A. charidemi* or that a deletion had occurred between *A. charidemi* and *A. majus*. However, the Tam661 sequences in *A. majus* and *A. charidemi* alleles showed no clear evidence for insertion or deletion causing the differences between alleles. However, *A. molle* carried no obvious sequence footprint and is therefore likely to represent the ancestral sequence before transposon insertion gave rise to the alleles in *A. majus* and *A. charidemi*. Transposon insertions upstream of coding sequences can change gene expression in a number of ways, either increasing or decreasing expression or altering spatial patterns of expression (reviewed by Weil and Wessler, 1990). It therefore seems possible that the transposon insertion in *Ros1* is responsible for differences in the activity between the *A. majus* and *A. charidemi* alleles. One possibility is that this creates a WRKY recognition site in the *A. majus* allele. However, there is no evident so far that WRKY can up-regulate genes that involving in anthocyanin biosynthesis.

Although differences in *Ros* expression were detected in plants carrying either *A. majus* or *A. charidemi Ros* alleles (Chapter 5), these experiments were not designed to further test the role of *Ros* expression in flower colour variation. For example, the differences in *Ros* RNA levels might have been caused by segregating linked or unlinked factors acting in *trans* on *Ros*. To do this, expression of the *Ros* alleles would have been better tested in a more isogenic background, for example in the progeny of a heterozygous NIL, or ideally in *ros* mutant plants transformed with the *Ros* allele from each of the species separately. It would also have been interesting to compare expression of the two alleles in corolla tubes and corolla lobes in these experiments. Because null *ros* mutants were not available and transforming *Antirrhinum* is time-consuming and not routine, these experiments were not pursued.

Other candidate genes for minor-effect QTL were also identified. Both *Candi* and *Del* were shown to be linked to a QTL in LG4 (at 12 cM) that affected colour variation in corolla tubes. *Candi* showed significant correlation with the corolla tube colour, therefore more-likely to be closely linked to the QTL. Given that *Del* is known to be

necessary for ABP gene expression in corolla tubes, it cannot be ruled out as a candidate for this QTL, though the correlation of *Del* genotype to corolla pigmentation was not significant. A minor QTL for lobe colour had been mapped around 58 cM in LG8 and a *WD40* 33 cM in the same chromosome, making it a candidate for this QTL.

Linkage maps are an important tool in genetic and evolution studies (e.g., Schwarz-Sommer *et al*., 2003). I was able to add several ABP genes to the linkage map for *Antirrhinum* (Table 7) that might be useful in future studies of anthocyanin variation in *Antirrhinum* and in assembling and annotating the *Antirrhinum* genome sequence.

Gene	Location		Reference	Note
	Chromosome	Position		
Nivea[1]	LG1	85cM	Feng *et al*., 2009	Structural gene
CHI[2]	LG4	62 cM	Schwarz-Sommer *et al*., 2003	Structural gene
Incolorata[1]	LG7	8cM	This experiment	Structural gene
Eosina	?	?	-	Structural gene
Pallida[2]	LG3	27cM	Schwarz-Sommer *et al*., 2003	Structural gene
Candica[1]	LG4	16cM	Feng *et al*., 2009	Structural gene
UFGT[2]	LG2	11cM	Schwarz-Sommer *et al*., 2003	Structural gene
FLS	?	?	-	Structural gene
Delila[1]	LG4	0cM	Feng *et al*., 2009	bHLH regulatory gene
Mutabilis[1]	LG4	56cM	This experiment	bHLH regulatory gene
WD40[1]	LG8	33cM	This experiment	WD repeat regulatory gene
Rosea1[1]	LG3	64cM	This experiment	MYB regulatory gene
Rosea2[1]	LG3	64cM*	This experiment*	MYB regulatory gene
Venosa [1]	LG6	53cM	This experiment	MYB regulatory gene

Table 7. The location of *Antirrhinum* ABP genes [1] = mapped in *A.majus* x *A.charidemi* F_2 population. [2] = mapped in *A.majus* x *A.molle* hybrids population. * Estimated location from a previous study (Schwinn *et al*., 2006).

The main focus of my study became the identification of a second major-effect QTL for flower colour in LG7. This especially affected lobe pigmentation, explaining 10% of the variance in the F_2 population (Table 5) and had been estimated to be around position 7cM. The only gene known to be involved in anthocyanin biosynthesis and located in LG7 was *Inc*, which was mapped to 8 cM in this study (Figure 16, Table 7). Because Inc activity appears to limit anthocyanin synthesis (i.e., *inc* loss-of-function mutations are semi-dominant because *inc*/+ heterozygotes have intermediate levels of anthocyanin), it had been assumed that *Inc* was the QTL. However, I was able to show in NILs that *Inc* was linked to the major-effect QTL but could be separated from it by recombination. The most concrete piece of evidences for this was that NIL family M239, segregated for the QTL but was homozygous for the *A. charidemi Inc* allele. Because no other candidate ABP genes mapped in this region, the LG7 QTL appeared to be novel and I concentrated on further attempts to identify its location and function.

Using gene-based molecular markers and AFLPs, I narrowed down the location of the LG7 QTL to a region of around 11 cM. The total length of the *A. majus* x *A. charidemi* linkage map is less than 700 cM (Feng *et al*., 2009) and the size of the *A. majus* genome is 634 Mb (Yong-Biao Xue, pers comm.), therefore 11 cM corresponds to an average of ~10 Mb of DNA. Unfortunately, the genome sequence of *A. majus* is not fully assembled and currently consists of several thousand sequence scaffold with an average size of ~0.1 Mb. The nearest gene-based markers to the LG 7 QTL (*ML-1* and *Inc*) were found to be located in different scaffold and therefore the sequence of the interval containing candidate genes for the QTL could not be identified within the time of this project.

The expression analysis of ABP genes in this project showed unexpected results in NILs generating from hybrids of *A. majus* and *A. charidemi*. Although these results were not confirmed using multiple replicates of different genotypes and quantified with q-RT-PCR, they strongly suggest that several ABP genes are up-regulated in pale flowers homozygous for the *A. charidemi* allele of the LG7 QTL. The simplest explanation for this effect was that ABP genes were up-regulated by a feedback response to decreased anthocyanin production or decreased levels of anthocyanin precursors, although this level

of control has yet to be demonstrated for the ABP. The findings therefore suggest that the regulation of anthocyanin pigmentation in flowers might be more complex than previously thought.

I have proposed two explanations for the effect of the LG7 QTL on anthocyanin levels, though both need further investigation because neither is fully supported by my experiments.

The first explanation is that the LG7 QTL involves an anthocyanidin modification gene (i.e *UFGT* or *3RT*) or an anthocyanin transporter. Therefore pale pigmentation could be the result of clogging of anthocyanin production by weak activity of the *A. charidemi* allele. Mutants in grape and Arabidopsis lacking the gene responsible for the initial glycosylation of cyanidin were reported to accumulate no anthocyanin, despite the presence of an intact anthocyanin biosynthesis pathway (Boss, Davies and Robinson, 1996; Tohge *et al*., 2005). This suggests that modification of anthocyanidin is required to stabilize the anthocyanin pigment, though these studies did not test the effects of mutations on expression of ABP genes. Because anthocyanin stability is decreased in mild acidic to neutral solutions (pH 5-7; Cabrita, Fossen and Anderson, 1999), another possibility is that reduced transport from the cytosol (pH ~ 7.2) to the vacuole (pH <5.0) might decrease anthocyanin levels. In either case, reduced anthocyanin levels might feed back onto structural gene expression in pale flowers and this might occur though increased expression of the regulators *Ros2* and *Del*.

The second explanation for the up-regulation of ABP structural genes was based on the observed pattern of structural gene expression. The most strongly up-regulated genes in pale flowers included *Niv*, *CHI*, *Inc* and *Eos*, similar to the genes up-regulated by R2R3 MYB subgroup 7 transcription factors (MYB11, MYB12 and MYB111) in Arabidopsis. This group of genes was reported to control early flavonoid biosynthesis pathway by targeting genes encoding Chalcone synthase (*CHS*), Chalcone isomerase (*CHI*), Flavonoid-3-hydroxylase (*F3H*), Flavonol synthase1 (*FLS1*) and flavonols:UDP-glucosyltransferase (*UGT91A1* and *UGT84A1*; Stracke *et al*., 2007) and also up-regulated

genes within the flavonol pathway. Therefore increased activity of a subgroup 7 MYB could increase ABP gene expression and divert more substrate from anthocyanin. However, the inconsistency of *CHI, Eos* and *FLS* expression in pale flowers argue against this, although *FLS* might not be the only member of this gene family involved in flavonol production in flowers (Chapter 5).

Preliminary tests of flavonoid and anthocyanin contents suggested that flavonoid:antocyanin ratios were higher in pale flowered plants, implying that the QTL affects flavonoid and anthocyanin production in different ways. However, without more accurate measurements of flavonols content, this does not reveal whether flavonoid synthesis is increased in pale flowers.

Even the introgression approach used in this project cannot overcome the major limitation in identifying genes underlying natural variation: that effects cannot strictly be ascribed to a single gene. However, my results so far suggest that the differences in floral anthocyanin pigmentation between *A. majus* and *A. charidemi* involves multiple genes that might act in different ways. Several reports suggest that natural variation in biosynthesis can be genetically complex, involving biosynthesis pathway, developmental regulation and epigenetic or allelic variation (Stracke *et al*., 2007; Durbin *et al*, 2003; Zufall and Rausher, 2004; Scwhinn *et al*., 2006; Whittall *et al*., 2006; Johannes *et al*., 2009; Quattrocchio *et al*., 1999). Mapping Antirrhinum QTLs has identified chromosome regions that make a major contribution to flower colour variation. This will help to identify candidate genes once the genome sequence of *A. majus* is fully assembled. Then candidate genes can be tested further, for example by introducing the parental alleles into null mutant backgrounds to measuring their effects on the phenotype.

Reference:

Aharoni, A., De Vos, C. H. R., Wein, M., Sun, Z. K., Greco, R., Kroon, A., Mol, J. N. M., and O'Connell, A. P. (2001). The strawberry *FaMYB1* transcription factor suppresses anthocyanin and flavonol accumulation in transgenic tobacco. Plant J 28:319-332.

Albert, N.W., Lewis, D.H., Zhang, H., Schwinn, K.E., Jameson, P.E., and Davies K.M. (2011) Members of an R2R3-MYB transcription factor family in Petunia are developmentally and environmentally regulated to control complex floral and vegetative pigmentation patterning. Plant J 65(5): 771–784.

Alfenito, M.R., Souer, E., Goodman, C.D., Buell, R., Mol, J., Koes, R., and Walbot, V. (1998). Functional complementation of anthocyanin sequestration in the vacuole by widely divergent glutathione S-transferases. Plant Cell 10: 1135–1149.

Almeida, J., Carpenter, R., Robin, T.P., Martin, C. and Coen, E.S. (1989) Genetics interaction underlying flower color pattern in *Antirrhinum majus*. Genes Dev 3: 1758-1767.

Altschul, S. F., Madden, T.L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller. W., and Lipman D.J.(1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25: 3389-3402.

Andersen, Ø. M., & Jordheim, M. (2006). The anthocyanins. In Ø. M. Andersen & K. R. Markham (Eds.), Flavonoids: Chemistry, Biochemistry, Applications. Boca Raton, FL, CRC Press: p 471–553.

Austin, M.B. and Noel, J.P. (2003) The chalcone synthase superfamily of type III polyketide synthases. Nat. Prod. Rep 20:79–110.

Bailey, P.C., Martin, C., Toledo-Ortiz, G., Quail, P.H., Huq, E., Heim, M.A., Jakoby, M., Werber, M., and Weisshaar, B. (2003) Update on the basic helix-loop-helix transcription factor gene family in *Arabidopsis thaliana*. Plant Cell. 15:2497–2502.

Bartlett, N.J.R. (1989) The genetic control of anthocyanin biosynthesis in *Antirrhinum majus*. Ph.D. Thesis, University of East Anglia.

Baudry, A., Caboche, M., and Lepiniec, L. (2006) *TT8* controls its own expression in a feedback regulation involving *TTG1* and homologous MYB and bHLH factors, allowing a strong and cell-specific accumulation of flavonoids in *Arabidopsis thaliana*. Plant J 46: 768–779.

Baur, E. (1910a). Vererbungs- und Bastardierungsversuche mit Antirrhinum. Z. Indukt. Abstammungs. Vererbungsl. 3: 34–98.

Baur, E. (1910b). Untersuchungen u¨ ber die Vererbung von Chromatophorenmerkmalen bei Melandrium, Antirrhinum und Aquilegia. Z. Indukt. Abstammungs. Vererbungsl. 4: 81–102.

Beld, M., Martin, C., Huits, H., Stuitje, A.R., and Gerats, A.G.M. (1989). Flavonoid synthesis in *Petunia hybrida*: Partial characterisation of dihydroflavonol 4-reductase genes. Plant Mol. Biol. 13: 491-502.

Bentham, J. (2001) Genographer 1.6. http://hordeum.oscs.montana.edu/genographer.

Bollmann, J., Carpenter, R., and Coen, E.S. (1991) Allelic lnteractions at the *nivea* Locus of *Antirrhinum*. Plant Cell 3:1327-1336.

Borevitz, J.O., Xia, Y., Blount, J., Dixon, R.A. and Lamb, C. (2000) Activation tagging identifies a conserved MYB regulator of phenylpropanoid biosynthesis. Plant Cell 12: 2383–2394.

Bonas, U., Sommer, H. and Saedler, H. (1984) The 17 kb Tam1 element of *Antirrhinum majus* induces a 3 bp duplication upon integration into chalcone synthase gene. EMBO J 3: 1015–1019.

Boss, P.K., Davies, C., and Robinson, S.P. (1996) Analysis of the expression of anthocyanin pathway genes in developing *Vitis vinifera* L. cv Shiraz grape berries and the implications for pathway regulation. Plant Physiol 111: 1059–1066.

Bradley, D., Carpenter, R., Copsey, L., Vincent, C., Rothstein, S., and Coen E. (1996) Control of inflorescence architecture in *Antirrhinum*. Nature 379: 791–797.

Bradshaw, H.D. and Schemske, D.W. (2003) Allele substitution at a flower color locus produces a pollinator shift in two monkeyflower species (*Mimulus*). Nature 426:176–178.

Britsch, L., Ruhnau-Brich, B., and Forkmann, G. (1992) Molecular cloning, sequence analysis, and in vitro expression of flavanone 3 beta-hydroxylase from *Petunia hybrida*. J Biol Chem 267(8): 5380–5387.

Brouillard, R. (1982) Chemical structure of anthocyanins. in: Anthocyanins as Food Color ed. Markakis, P., Academic Press, New York: 1–40.

Broun, P. (2005). Transcriptional control of flavonoid biosynthesis: a complex network of conserved regulators involved in multiple aspects of differentiation in Arabidopsis. Curr. Opin. Plant Biol 8: 272–279.

Brugliera, F., Holton, T.A., Stevenson, T.W., Farcy, E., Lu, C.Y., and Cornish. E.C. (1994) Isolation and characterization of a cDNA clone corresponding to the *Rt* locus of *Petunia hybrida*. Plant J 5(1):81–92.

Brugliera, F., Tull D, Holton, T.A., Karan, M., Treloar, N., Simpson, K., Skurczynska, J., and Mason, J.G. (2000) Introduction of a cytochrome *b5* enhances the activity of flavonoid 3'5' hydroxylase (a cytochrome P450) in transgenic carnation. Sixth International Congress of Plant Molecular Biology. University of Laval, Quebec, S6–S8.

Bongue-Bartelsman, M., O'Neill, S.D., Tong, Y. and Yoder, J. I. (1994) Characterization of the gene encoding dihydroflavonol 4-reductase in tomato. Gene 138: 153–157.

Butelli, E., Titta, L., Giorgio, M., Mock, H ., Matros, A., Peterek, S., Schijlen, E.G.W.M., Hall, R.D., Bovy, A.G., Luo, J.. and Martin, C. (2008) Enrichment of tomato fruit with health-promoting anthocyanins by expression of select transcription factors. Nature Biotech 26 (11): 1301-1308.

Cabrita, L., Fossen, T., and Andersen, Ø.M. (2000) Colour and stability of the six common anthocyanidin 3-glucosides in aqueous solutions. Food Chem 68:101–107.

Chaffe, C.J. (2002) Molecular and Genetic Characterisation of European Snapdragons (Scrophulariaceae: *Antirrhinum* spp.). PhD thesis, University of East Anglia.

Chandler, V.L., Radicella, J.P., Robbins, T.P., Chen, J., and Turks, D. (1989) Two regulatory genes of the maize anthocyanin pathway are homologous: isolation of *B* utilizing *R* genomic sequences. Plant Cell 1: 1175–1183.

Chatel, G., Montiel, G., Pré, M., Memelink, J., Thiersault, M., SaintPierre, B., Doireau, P., and Gantet, P. (2003) *CrMYC1*, a *Catharanthus roseus* elicitor- and jasmonate-responsive bHLH transcription factor that binds the G-box element of the strictosidine synthase gene promoter. J Exp Bot 54: 2587–2588.

Chen, M., SanMiguel, P., Bennetzen, J. L. (1998) Sequence organization and conservation in *sh2/a1*-homologous regions of sorghum and rice. Genetics 148: 435–443.

Chittka, L. and Waser N.M. (1997) Why red flowers are not invisible for bees. Isr J Plant Sci 45:169–183.

Ciolkowski, I., Wanke, D., Birkenbihl, R.P., and Somssich, I.E.(2008) Studies on DNA-binding selectivity of WRKY transcription factors lend structural clues into WRKY-domain function. Plant Mol Biol 68: 81–92.

Clegg, M.T., and Durbin, M.L. (2000) Flower color variation: A model for the experimental study of evolution. PNAS 97(13):7016–7023.

Clegg, M.T., and Durbin, M.L. (2003) Tracing Floral Adaptations from Ecology to Molecules. Nature Genetics 4: 206-215.

Coberly, L.C. and Rausher, M.D. (2008) Pleiotropic effects of an allele producing white flowers in *Ipomoea purpurea*. Evolution 62:1076–1085.

Coen, E.S., Carpenter, R., and Martin, C. (1986) Transposable elements generate novel spatial patterns of gene expression in *Antirrhinum majus.* Cell 47: 285-296.

Coen, E. S., Romero, J. M., Doyle, S., Elliott, R., Murphy, G., and Carpenter, R. (1990) *FLORICAULA*: a homeotic gene required for flower development in *Antirrhinum majus*. Cell 63:1311–1322.

Colot, V., Haedens, V., and Rossignol, J. (1998) Extensive, Nonrandom Diversity of Excision Footprints Generated by *Ds*-Like Transposon *Ascot-1* Suggests New Parallels with V(D)J Recombination. Mol Cell Biol 18(7): 4337–4346.

Comba, L., Corbet, S.A., Hunt, H., Outram, S., Parker, J.S., and Glover, B.J. (2000) The role of genes influencing the corolla in pollination of *Antirrhinum majus*. Plant Cell Environ 23:639–647.

Cubas, P.,Vincent, C., and Coen, E. (1999) An epigenetic mutation responsible for natural variation in foral symmetry. Nature 401: 157-161.

Czemmel, S., Stracke, R., Weisshaar, B., Cordon, N., Harris, N.N., Walker, A.R., Robinson, S.P., and Bogs, J. (2009) The grapevine R2R3-MYB transcription factor VvMYBF1 regulates flavonol synthesis in developing grape berries. Plant Physiol 151: 1513–1530.

Darwin, C.R. (1868) The Variation of Animals and Plants under Domestication. Volume 2. London: John Murray.

de Vetten, N., Quattrocchio, F., Mol, J., and Koes, R. (1997) The an11 locus controlling flower pigmentation in petunia encodes a novel WD-repeat protein conserved in yeast, plants, and animals. Genes and Development 11: 1422–1434.

de Vetten, N., ter Horst, J., van Schaik, H.P., de Boer, A.J.M., Mol, J. and Koes, R. (1999) A cytochrome *b5* is required for full activity of flavonoid 3',5'hydroxylase, a cytochrome P450 involved in the formation of blue flower colors. Proc Natl Acad Sci USA 96: 778–783.

Demarest, K., Koyner, J., McCaughran, J. Jr., Cipp, L., Hitzemann, R. (2001) Further characterization and high-resolution mapping of quantitative trait loci for ethanol-induced locomotor activity. Behav Genet 31: 79–91.

Dixon, R.A., Blyden, E.R., Robbins M.P., van Tunen, A.J., and Mol J.N.M (1988) Comparative biochemistry of Chalcone isomerase. Phytochemistry 27 (9): 2801-2808.

Doyle, J.J. and Doyle, J.L. (1990) Isolation of plant DNA from fresh tissue. Focus 12(1): 13-15.

Dubos, C., Stracke, R., Grotewold, E., Weisshaar, B., Martin, C. and Lepiniec, L. (2010) MYB transcription factors in Arabidopsis. Trends Plant Sci. 15: 573–581.

Durbin, M.L., Learn J.R., G.H., Huttley, G.A. and Clegg, M.T. (2000a) Evolution of the chalcone synthase gene family in the genus Ipomoea. PNAS 92: 3338-3342.

Durbin, M.L., McCaig, B. and Clegg, M.T. (2000b) Molecular evolution of the chalcone synthase multigene family in the morning glory genome. Plant Mol. Biol. 42:79–92.

Durbin, M.L., Lundy, K.E., Morrell, P.L., Torres-Martinez, C.L. and Clegg, M.T. (2003) Genes that determine flower color: the role of regulatory changes in the evolution of phenotypic adaptations. Mol Phylogenetics Evol 29: 507–518.

Dyer, A.G., Whitney, H.M., Arnold, S.E.J., Glover, B.J., and Chittka, L. (2006) Bees associate warmth with floral colour. Nature 442:525.

Dyer, A.G., Whitney, H.M., Arnold, S.E.J., Glover, B.J., and Chittka L. (2007) Mutations perturbing petal cell shape and anthocyanin synthesis influence bumblebee perception of *Antirrhinum majus* flower colour. Arthropod-Plant Interactions 1:45–55.

Edwards, K., Johnstone, C. and Thompson, C. (1991) A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. Nucl. Acid. Res. 19: 1349.

Feller, A., Machemer, K., Braun, E.L. and Grotewold, E. (2011) Evolutionary and comparative analysis of MYB and bHLH plant transcription factors. Plant J 66(1): 94–116.

Feng, X., Wilson, Y., Bowers, J., Kennaway, R., Bangham, A., Hannah, A., Coen, E. and Hudson, A. (2009). Evolution of allometry in *Antirrhinum*. Plant Cell 21:2999-3007.

Ferreyra, M.L., Rius, S., Emiliani, J., Pourcel, L., Feller, A., Morohashi, K., Casati, P. and Grotewold, E. (2010). Cloning and characterization of a *UV-B*-inducible maize flavonol synthase. Plant J 62: 77–91.

Frangne, N., Eggmann, T., Koblischke, C., Weissenböck, G., Martinoia, E. and Klein, M. (2002) Flavone Glucoside Uptake into Barley Mesophyll and Arabidopsis Cell Culture Vacuoles; Energization Occurs by H^+-Antiport and ATP-Binding Cassette-Type Mechanisms. Plant Physiol 128: 726–733.

Froemel, S., de Vlaming, P., Stotz, G., Wiering, H., Forkmann, G. and Schram, A.W. (1985): Genetic and biochemical studies on the conversion of flavanones to dihydroflavonols in flowers of *Petunia hybrida*. Theor Appl Genet 70:561-568.

Forkmann, G. and Heller, W. (1999) Biosynthesis of flavonoids. In: Sankawa U (ed) Polyketides and Other Secondary Metabolites Including Fatty Acid and Other Derivatives. Elsevier, Amsterdam, p 713–748.

Filla, A., DeMichele, G., Cavalcanti, F., Pianese, L., Monticelli, A., Campanella, G. and Cocozza, S. (1996) The relationship between trinucleotide (GAA) repeat length and clinical features in Friedreich ataxia. Am. J. Hum. Genet. 59:554–60.

Fujiwara, H., Tanaka, Y., Yonekura-Sakakibara, K., Fukuchi-Mizutani, M., Nakao, M., Fukui, Y., Yamaguchi, M., Ashikari, T. and Kusumi, T. (1998) cDNA cloning, gene expression and subcellular localization of anthocyanin 5-aromatic acyltransferase from *Gentiana triflora*. Plant J 16:421-431.

Fukuchi-Mizutani, M., Okuhara, H., Fukui, Y., Nakao, M., Katsumoto, Y., Yonekura-Sakakibra, K., Kusumi, T., Hase, T. and Tanaka, Y. Biochemical and molecular characterization of a novel UDP-glucose: Anthocyanin 3'-O-glucosyltransferase, a key enzyme for blue anthocyanin biosynthesis, from gentian. Plant J. 2003, 132: 1652–1663.

Gachon, C.M., Langlois-Meurine, M., and Sainrenan, P. (2005) Plant secondary metabolism glycosyltransferases: the emerging functional analysis. Trends Plant Sci 10: 542–549.

Gattuso, G., Barreca, D., Gargiulli, C., Leuzzi, U. and Caristi, C. (2007) Flavonoid Composition of Citrus Juices Molecules 12, 1641-1673.

Gemayel, R., Vinces, M.D., Legendre, M. and Verstrepen, K.J. (2010) Variable Tandem Repeats Accelerate Evolution of Coding and Regulatory Sequences. Annu. Rev. Genet 44: 445-477.

Glover, B. J. and Martin, C. (1998). The role of petal cell shape and pigmentation in pollination success in *Antirrhinum majus*. Heredity 80, 778- 784.

Glover, B. J., Perez-Rodriguez, M. and Martin, C. (1998). Development of several epidermal cell types can be specified by the same MYB-related plant transcription factor. Development 125, 3497-3508

Gomez, C., Terrier, N., Torregrosa, L., Vialet, S., Fournier-Level, A., Verriès, C., Souquet, J., Mazauric, J., Klein, M., Cheynier, V. and Ageorges, A. (2009) Grapevine MATE-Type Proteins Act as Vacuolar H^+-Dependent Acylated Anthocyanin Transporters. Plant Physiol 150: 402–415.

Gong, Z.Z., Yamagish,i E., Yamazaki, M. and Saito, K. (1999) A constitutively expressed Myc-like gene involved in anthocyanin biosynthesis from *Perilla frutescens*: molecular characterization, heterologous expression in transgenic plants and transactivation in yeast cells. Plant Mol Biol 41: 33–44.

Goodman, C.D., Casati, P. and Walbot, V. (2004) A multidrug resistance-associated protein involved in anthocyanin transport in *Zea mays*. Plant Cell 16: 1812-1826.

Goodrich, J., Carpenter, R. and Coen, E.S. (1992). A common gene regulates pigmentation pattern in diverse plant species. Cell 68, 955-964.

Greenshields, D. (2007) Isolation of adaptive quantitative trail loci in *Antirrhinum.* PhD Thesis, University of Edinburgh.

Grotewold, E. and Davies, K. (2008) Trafficking and Sequestration of Anthocyanins. Nat Prod Commun 3: 1251-1258.

Grotewold, E., Chamberlain, M., St. Claire, G., Swenson, J., Siame, B.A., Butler, L.G., Snook, M. and Bowen, B. (1998) Engineering secondary metabolism in maize cells by ectopic expression of transcription factors. Plant Cell 10:721–740.

Haley, C.S., Knott, S.A. and Elsen, J. (1994) Mapping Quantitative Trait Loci in Crosses Between Outbred Lines Using Least Squares. Genetics 136:1195-1207.

Harborne, J. B. (1963) Plant polyphenols 10. Flavone and aurone glycosides of Antirrhinum. Phytochemistry 2:327-334.

Hatayama, M., Ono, E.,Yonekura-Sakakibara, K.,Tanaka, Y., Nishino, T. and Nakayama, T. (2006) Biochemical characterization and mutational studies of a chalcone synthase from yellow snapdragon (*Antirrhinum majus*) flowers. Plant Biotech 23: 373–378.

Helariutta, Y. , Elomaa, P., Kotilainen, M., Seppäinen, P. and Teeri ,T.H. (1993) Cloning of cDNA coding for dihydroflavonol-4-reductase (DFR) and characterization of *dfr* expression in the corollas of *Gerbera hybrida* var. Regina (Compositae). Plant Molecular Biology 22: 183-193.

Heller, W., and Forkmann, G. (1988). Biosynthesis. In The Flavonoids, J.B. Harbone, ed., London: Chapman and Hall, p 399-425.

Hichri, I., Barrieu, F., Bogs, J., Kappel, C., Delrot, S., and Lauvergeat, V. (2011) Recent advances in the transcriptional regulation of the flavonoid biosynthetic pathway. J. Exp. Bot. 62 (8): 2465-2483.

Hoballah, M.E., Gubitz, T., Stuurman ,J., Broger, L., Barone, L., Mandel, T., Dell'Olivo. A., Arnold, M. and Kuhlemeier, C. (2007) Single gene-mediated shift in pollinator attraction in *Petunia*. Plant Cell. 19:779–790.

Holton, T.A. and Cornish, E.C.(1995) Genetics and biochemistry of anthocyanin biosynthesis. Plant Cell 7, 1071-1083.

Hopkins, R. and Rausher, M.D.(2011) Identification of two genes causing reinforcement in the Texas wildflower *Phlox drummondii*. Nature. 469(7330): 411-414.

Hopp, W. and Seitz, H. (1987) The uptake of acylated anthocyanin into isolated vacuoles from a cell suspension culture of *Daucus carota*. Planta 170:74-85.

Hoshino, A., Abe, Y., Saito, N., Inagaki, Y. and Iida, S., (1997) The gene encoding flavanone 3- hydroxylase is expressed normally in the pale yellow flowers of the Japanese morning glory carrying the speckled mutation which produce neither flavonol nor anthocyanin but accumulate chalcone, aurone and flavanone, Plant Cell Physiol 38: 970-974.

Hoshino, A., Morita, Y., Choi, J.D., Saito, N., Toki, K., Tanaka, Y., and Iida, S. (2003) Spontaneous Mutations of the Flavonoid 3'-hydroxylase Gene Conferring Reddish Flowers in the Three Morning Glory Species. Plant Cell Physiol. 44(10): 990–1001

Howard, J. (2009) Why didn't Darwin discover Mendel's laws? J Biol 8:15.

Hudson, A., Critchley, J. and Erasmus Y. (2008a) The genus Antirrhinum (snapdragon): A flowering plant model for evolution and development. Cold Spring Harb. Protoc. doi:10.1101/pdb.emo100

Hudson A., Critchley J., and Erasmus Y. (2008b) Cultivating *Antirrhinum*. Cold Spring Harb. Protoc. doi:10.1101/pdb.prot5051

Huits, H.S.M., Gerats, A.G.M., Kreike, M.M., Mol, J.N.M. and Koes, R.E. (1994). Genetic control of dihydroflavonol 4-reductase gene expression in *Petunia hybrida*. Plant J 6: 295–310.

Inagaki, Y., Hisatomi, Y., Suzuki, T., Kasahara, K. and Iida, S. (1994) Isolation of a Suppressormutator/Enhancer-like transposable element, *Tpn1*, from Japanese morning glory bearing variegated flowers. Plant Cell 6: 375-383.

Inagaki, Y., Johzuka-Hisatomi, Y., Mori, T., Takahashi, S., Hayakawa, Y., Peyachoknagul, S., Ozeki Y., and Iida, S. (1999) Genomic organization of the genes encoding dihydroflavonol 4-reductase for flower pigmentation in the Japanese and common morning glories. Gene 226:181–188.

Jackson, D., Roberts, K., and Martin, C. (1992). Temporal and spatia1 control of expression of anthocyanin biosynthetic genes in developing flowers of *Antirrhinum majus*. Plant J. 2, 425-434.

Johannes, F., Porcher, E., Teixeira, F.K, Saliba-Colombani, V., Simon, M., Agier, N., Bulski, A. X., Albuisson, J., Heredia, F., Audigier, P., Bouchez, D., Dillmann, C., Guerche, P., Hospital, F., and Colot V. (2009) Assessing the Impact of

Transgenerational Epigenetic Variation on Complex Traits. PLoS Genet 5(6): e1000530. doi:10.1371/journal.pgen.1000530

Johnson, E.T., Ryu, S., Yi, H., Shin, B., Cheong, H,. and Choi, G. (2001) Alteration of a single amino acid changes the substrate specificity of dihydroflavonol 4-reductase. Plant J 25: 325-333.

Johzuka-Hisatomi, Y., Hoshino, A., Mori, T., Habu, Y., and Iida, S. (1999) Characterization of the chalcone synthase genes expressed in flowers of the common and Japanese morning glories. Genes Genet Syst.74(4):141-147.

Jones, K.N. and Reithel, J.S. (2001) Pollinator-mediated selection on a flower color polymorhism in experimental populations of *Antirrhinum* (Scrophulariaceae). Amer J of Bot 88(3): 447-454.

Jones, P., Messner, B., Nakajima, J., Schäffner, A.R., and Saito, K. (2003) *UGT73C6* and *UGT78D1*, Glycosyltransferases Involved in Flavonol Glycoside Biosynthesis in Arabidopsis thaliana. J Biol Chem 278 (45): 43910–43918.

Jong, A.Y., T'ang, A., Liu, D.P., and Huang, S.H. (2002) Inverse PCR. Genomic DNA cloning. Methods Mol Biol 192:301-307.

Jonsson, L.M.V., Aarsman, M.E.G., Poulton, J.E., and Schram, A.W.(1984a) Properties and genetic control of four methyltransferases involved in methylation of anthocyanins in flowers of *Petunia hybrida*. Planta 160:171-179.

Jonsson, L.M.V., Aarsman, M.E.G., van Diepe n, J., de Vlaming, P., Smit, N., and Schram, A.W.(1984b) Properties and genetic control of anthocyanin 5-O-glucosyltransferase in flowers of *Petunia hybrida*. Planta 160:341-347.

Kevan, P., Glurfa, M. and Chittka, L. (1996) Why are there so many and so few white flowers? Trends Plant Sci 1(8): 280-284.

Kevan, P.G., Chittka, L. and Dyer, A.G. (2001) Limits to the salience of ultraviolet: lessons from colour vision in bees and birds. J Exp Biol 204:2571–2580.

Kim, C.Y., Lee, H.J., Lee, E.H., Jung, S.H., Lee, D.U., Kang, S.W., Hong, S.J., and Um, B.H. (2008). Rapid identification of radical scavenging compounds in blueberry extract by HPLC coupled to an on-line ABTS based assay and HPLC-ESI/MS. Food Sci. Biotechnol., 17: 846-849.

Klein, M., Weissenbock, G., Dufaud, A., Gaillard, C., Kreuz, K., and Martinoia, E. (1996) Different energization mechanisms drive the vacuolar uptake of a flavonoid glucoside and a herbiside glucoside. J Biol Chem 271:29666-29671.

Koes, R., Verweij, W., and Quattrocchio, F. (2005) Flavonoids: a colorful model for the regulation and evolution of biochemical pathways. Trends Plant Sci 10(5): 236-242.

Koes, R.E, Spelt, C.E., van den Elzen, P.J.M, Mol, J.N.M (1989) Cloning and molecular characterization of the chalcone synthase multigene family of *Petunia hybrida.* Gene 81(2): 245–257.

Koes, R.E., Spelt, C.E., Reif, H.J., van den Elzen, P.J.M., Veltkamp E., and Mol, J.N.M. (1986) Floral tissue of *Petunia hybrida* (V30) expresses only one member of the chalcone synthase multigene family. Nucl Acids Res 14 (13): 5229-5239.

Kolosova, N., Gorenstein, N., Kish, C.M., and Dudareva, N. (2001) Regulation of circadian methyl benzoate emission in diurnally and nocturnally emitting plants. Plant Cell 13:2333-2347.

Kosambi, D.D. (1944) The estimation of map distance from recombination values. Ann. Eugen12: 172-175.

Kristiansen KN, Rohde W (1991) Structure of the Hordeum vulgare gene encoding dihydroflavonol 4-reductase and molecular analysis of *ant18* mutants blocked in flavonoid synthesis. Mol Gen Genet 230: 49–59.

Kroon, A.R. (2004). Transcription Regulation of the Anthocyanin Pathway inPetunia hybrida. PhD dissertation (Amsterdam: Vrije Universiteit).

Kroon, J., Souer, E., de Graaff, A., Xue, Y., Mol. J., and Koes, R. (1994) Cloning and structural analysis of the anthocyanin pigmentation locus *Rt* of *Petunia hybrida*: characterization of insertion sequences in two mutant alleles. Plant J 5(1):69–80.

Kuckuck, H., and Schick, R. (1930). Die Erbfaktoren bei Antirrhinum majus und ihre Bezeichnung. Z. Indukt. Abstammungs. Vererbungsl.56, 51–83.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettiga,n P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., and Higgins, D.G. (2007) ClustalW and ClustalX version 2. Bioinformatics 23(21): 2947-2948.

Legare, M.E. and Frankel, W.N. (2000) Multiple seizure susceptibility genes on chromosome 7 in SWXL-4 congenic mouse strains. Genomics 70: 62–65.

Lepiniec, L., Debeaujon, I., Routaboul, J.-M., Baudry, A., Pourcel, L., Nesi, N., and Caboche, M. (2006). Genetics and biochemistry of seed flavonoids. Annu. Rev. Plant Biol. 57: 405–430.

Li, X., Duan, X., Jiang, H., Sun, Y., Tang, Y., Yuan, Z., Guo, J., Liang, W., Chen, L., Yin, J., Ma, H., Wang, J. and Zhang, D. (2006) Genome-Wide Analysis of Basic/Helix-Loop-Helix Transcription Factor Family in Rice and Arabidopsis. Plant Physiol 141(4): 1167-1184.

Lister, C., Jackson, D., and Martin, C. (1993) Transposon-induced inversion in *Antirrhinum* modifies *nivea* gene expression to give a novel flower color pattern under the control of cycloidearadialis. Plant Cell 5: 1541–1553.

Lowry, D.B, Sheng, C.C., Lasky, J.R., and Willis, J.H. (2012) Five anthocyanin polymorphisms are associated with an R2R3-MYB cluster in *Mimulus guttagus* (Phrymaceae). Amer J Bot 99(1): 82–91.

Lu, Y., Du, J., Tang, J., Wang, F., Zhang, J., Huang, J., Liang, W. and Wang, L. (2009) Environmental regulation of floral anthocyanin synthesis in *Ipomoea purpurea*. Mol Ecol 18: 3857–3871.

Luo, D., Carpenter, R., Copsey, L., Vincent, C., Clark, J. and Coen, E.S. (1999). Control of organ asymmetry in flowers of *Antirrhinum*. Cell 99: 367–376.

Luo, D., Coen, E.S., Doyle, S. and Carpenter, R. (1991) Pigmentation mutants produced by transposon mutagenesis in *Antirrhinum majus*. Plant J 1(1): 59-69.

Marais, D.L.D. and Rausher, M.D. (2010) Parallel evolution at multiple levels in the origin of hummingbird pollinated flowers in *Ipomoea*. Evolution 64(7): 2044–2054.

Marinova, K., Pourcel, L., Weder, B., Schwarz, M., Barron, D., Routaboul, J.M., Debeaujon, I. and Klein, M. (2007) The Arabidopsis MATE transporter TT12 acts as a vacuolar flavonoid/H+-antiporter active in proanthocyanidin-accumulating cells of the seed coat. Plant Cell 19: 2023–2038.

Marrs, K.A., Alfenito, M.R., Lloyd, A.M. and Walbot, V. (1995) A glutathione S-transferase involved in vacuolar transfer encoded by the maize gene *Bronze-2*. Nature 375: 397-400.

Martin, C., Prescott, A., Machay, S., Bartlett, J. and Vrijklandt, E. (1991) The control of anthocyanin biosynthesis in flowers of *Antirrhinum majus*. Plant J 1: 37-49.

Martin, C. and Gerats, T. (1993) Control of Pigment Biosynthesis Genes during Petal Development. Plant Cell 5: 1253-1264.

Martin, C. and Paz-Ares, J. (1997) MYB transcription factors in plants. Trends Genet 13: 67–73.

Martin, C., Carpenter, R., Sommer, H., Saedler, H. and Coen, E.S. (1985) Molecular analysis of instability in flower pigmentation of *Antirrhinum majus*, following isolation of the *pallida* locus by transposon tagging. EMBO 4(7):1625–1630.

Matern, U., Heller, W., and Himmelspach, K. (1983) Conformational Changes of apigenin 7-0-(6-0-malonylglucoside), a Vacuolar Pigment from Parsley, with Solvent Composition and Proton Concentration. Eur. J. Biochem. 133: 439-448.

Mathews, H., Clendennen, S.K., Caldwell, C.G., Liu, X.L., Connors, K., Matheis, N., Schuster, D.K., Menasco, D.J., Wagoner, W., Lightner, J. and Wagner, D. (2003) Activation tagging in tomato identifies a transcriptional regulator of anthocyanin biosynthesis, modification, and transport. Plant Cell 15 (8): 1689-1703.

Massari, M. E., Murre C. (2000) Helix-Loop-Helix Proteins: Regulators of Transcription in Eucaryotic Organisms. Mol Cell Biol 20: 429-440.

McCaig, B. (1998) The molecular evolution and expression of anthocyanin multigene families in *Ipomoea purpurea* (common morning glory): pathway evolution. Ph.D. Dissertation, University of California, Riverside.

Morita, Y., Hoshino, A., Kikuchi, Y., Okuhara, H., Ono, E., Tanaka, Y., Fukui, Y., Saito, N., Nitasaka, E., Hiroshi Noguchi, H. and Iida, S. (2005) Japanese morning glory dusky mutants displaying reddish-brown or purplish-gray flowers are deficient in a novel glycosylation enzyme for anthocyanin biosynthesis, UDP-glucose:anthocyanidin 3-O-glucoside-2''-O-glucosyltransferase, due to 4-bp insertions in the gene. Plant J 42(3):353-63.

Morita, Y., Saitoh, M., Hoshino, A., Nitasaka, E. and Iida, S. (2006) Isolation of cDNAs for R2R3-MYB, bHLH and WDR transcriptional regulators and identification of *c* and *ca* mutations conferring white flowers in the Japanese morning glory. Plant Cell Physiol 47:457–470.

Mozhui, K., Ciobanu, D.C., Schikorski, T., Wang, X., Lu, L. and Williams, W.W. (2008) Dissection of a QTL Hotspot on Mouse Distal Chromosome 1 that Modulates Neurobehavioral Phenotypes and Gene Expression. PLoS Genet 4(11): e1000260. doi:10.1371/journal.pgen.1000260

Mueller, L.A., Goodman, C.D., Silady, R.A. and Walbot, V. (2000) AN9, a petunia glutathione S-transferase required for anthocyanin sequestration, is a flavonoid-binding protein. Plant Physiol 123(4):1561-70.

Nakajima, J., Tanaka, Y., Yamazaki, M., and Saito, K. (2001) Reaction Mechanism from Leucoanthocyanidin to Anthocyanidin 3-Glucoside, a Key Reaction for Coloring in Anthocyanin Biosynthesis. J Biol Chem 276(28): 25797–25803.

Negre, F., Kish, C.M., Boatright, J., Underwood, B., Shibuya, K., Wagner, C., Clark, D.G., and Dudareva, N. (2003) Regulation of methylbenzoate emission after pollination in snapdragon and petunia flowers. Plant Cell 15:2992-3006.

Nesi, N., Jond, C., Debeaujon, I., Caboche, M. and Lepiniec, L. (2001) The Arabidopsis *TT2* gene encodes an R2R3 MYB domain protein that acts as a key determinant for proanthocyanidin accumulation in developing seeds. Plant Cell 13: 2099–2114.

Noda, K., Glover, B.J., Linstead, P., and Martin, C. (1994) Flower colour intensity depends on specialised cell shape controlled by a MYB-related transcription factor. Nature 369:661–664.

Ogata, J., Kanno, Y., Itoh, Y., Tsugawa, H. and Suzuki, M. (2005) Plant biochemistry: Anthocyanin biosynthesis in Roses. Nature 435, 757-758.

Olmstead, R. G., DePamphilis,C. W., Wolfe, A. D., Young, N. D., Elisons, W. J. and Reeves, P. A. (2001) Disintegration of the Scrophulariaceae. Amer J Bot 88:348-361.

Pathirana, N.N. (2007) Mechanism of complex programmed patterns of anthocyanin pigment formation in *Antirrhinum majus*. PhD thesis, Massey University, New Zealand.

Peitsch, D., Fietz, A., Hertel, H., de Souza, J., Ventura, D.F. and Menzel, R (1992) The spectral input systems of hymenopteran insects and their receptor-based colour vision. J Comp Physiol A 170:23–40.

Perez-Rodriguez, M., Jaffe, F. W., Butelli, E., Glover, B. J. and Martin, C. (2005). Development of three different cell types is associated with the activity of a specific MYB transcription factor in the ventral petal of *Antirrhinum majus* flowers. Development 132: 359-370.

Petkov, P.M., Graber, J.H.. Churchill, G.A., DiPetrillo, K., King, B.L., and Paigen, K. (2006) Evidence of a large-scale functional organization of mammalian chromosomes. PLoS Genet 1: e33.

Pires , N. and Dolan, L. (2010) Origin and Diversification of Basic-Helix-Loop-Helix Proteins in Plants. Mol Biol Evol 27 (4): 862-874.

Preuβ, A., Stracke, R., Weisshaar, B., Hillebrecht, A., Matern, U. and Martens, S. (2009) *Arabidopsis thaliana* expresses a second functional flavonol synthase. FEBS Letters 583:1981–1986.

Quattrocchio, F., Verweij, W., Kroon, A., Spelt, C., Mol, J. and Koes, R. (2006). PH4 of Petunia is an R2R3 MYB protein that activates vacuolar acidification through interactions with basic-helix-loop-helix transcription factors of the anthocyanin pathway. Plant Cell. 18: 1274–1291.

Quattrocchio, F., Wing, J.F., Leppen, H.T., Mol, J.N.M. and Koes, R. (1993) Regulatory genes controlling anthocyanin pigmentation are functionally conserved among plant species and have distinct sets of target genes. Plant Cell 5: 1497–1512.

Quattrocchio, F., Wing, J., van der Woude, K., Souer, E., de Vetten, N., Mol, J. and Koes, R. (1999) Molecular analysis of the *anthocyanin2* gene of petunia and its role in the evolution of flower color. Plant Cell 11: 1433–1444.

Ralston, L., Subramanian, S., Matsuno, M. and Yu, O. (2005) Partial reconstruction of flavonoid and isoflavonoid biosynthesis in yeast using soybean type I and type II chalcone isomerases. Plant Physiol 137:1375–1388.

Ramsay, N.A. and Glover, B.J. (2005). MYB-bHLH-WD40 protein complex and the evolution of cellular diversity. Trends Plant Sci 10: 63–70.

Rothmaler W. (1956) Taxonomische Monographie der Gattung Antirrhinum. Akademie-Verlag, Berlin, Germany.

Saito, K., Kobayashi, M., Gong, Z.Z., Tanaka, Y. and Yamazaki, M. (1999) Direct evidence for anthocyanidin synthase as a 2-oxoglutarate-dependent oxygenase: molecular cloning and functional expression of cDNA from a red forma of *Perilla frutescens*. Plant J. 17: 181–190.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, NY, Vol. 1, 2, 3.

Schwarz-Sommer, Z, Davies,B. and Hudson A. (2003) An everlasting pioneer: the story of *Antirrhinum* research. Nat Rev Genet 4, 655-663.

Schwarz-Sommer, Z., de Andrade Silva, E., Berndtgen, R., Lönnig, W., Müller, A., Nindl, I., Stüber, K., Wunder, J., Saedler,H., Gübitz, T., Borking, A.,Golz, J.F., Ritter, E. and Hudson, A. (2003) A Linkage Map of an F2 Hybrid Population of *Antirrhinum majus* and *A. molle*. Genetics 163: 699–710.

Schwarz-Sommer, Z., Gubitz, T., Weiss, J., Gomez-di-Marco, P., Delgado-Benarroch, L., Hudson, A., and Egea-Cortines, M. (2010). A molecular recombination map of *Antirrhinum majus*. BMC. Plant Biol. 10: 275-86.

Schwinn, K., Davies, K., Alm, V., Mackay, V. and Martin, C. (2001). Regulation of anthocyanin biosynthesis in *Antirrhinum*. Acta Hort. (ISHS) 560:201-206 http://www.actahort.org/books/560/560_37.htm

Schwinn, K., Venail, J., Shang, Y., Mackay, S., Alm, V., Butelli, E., Oyama, R., Bailey, P., Davies, K. and Martin, C. (2006). A small family of MYB-regulatory genes controls floral pigmentation intensity and patterning in the genus Antirrhinum. Plant Cell. 18, 831–851.

Seaton, G., Hernandez, J., Grunchec, J.A., White, I., Allen, J., De Koning, D.J., Wei, W., Berry, D., Haley, C. and Knott, S. (2006) GridQTL: A Grid Portal for QTL Mapping of Compute Intensive Datasets. Proceedings of the 8th World Congress on Genetics Applied to Livestock Production, August 13-18, 2006 Belo Horizonte, Brazil 2006.

Shang, Y. (2006) How the Pigment Stripes Form in Snapdragon (*Antirrhinum majus*) Flowers: a study of the molecular mechanism of venation pigmentation patterning in flowers. PhD thesis, Massey University, New Zealand.

Shang, Y., Venail, J., Mackay, S., Bailey, P.C., Schwinn, K.E., Jameson, P.E., Martin, C.R. and Davies, K.M. (2011) The molecular basis for venation patterning of pigmentation and its effect on pollinator attraction in flowers of *Antirrhinum*. New Phytol 189(2):602-15.

Shimada, N., Aoki, T., Sato, S., Nakamura, Y., Tabata, S. and Ayabe, S. (2003) A cluster of genes encodes the two types of chalcone isomerase involved in the biosynthesis of general flavonoids and legume-specific 5-deoxy (iso) flavonoids in *Lotus japonicus*. Plant Physiol 131:941–951.

Shimada, S., Otsuki, H. and Sakuta M. (2006) Transcriptional control of anthocyanin biosynthetic genes in the Caryophyllales. J Exp Bot 58: 957–967.

Shirley, B.W., Kubasek, W.L., Stroz, G., Bruggemann, E., Koornneef, M., Ausubel, F.M., and Goodman, H.M. (1995) Analysis of *Arabidopsis* mutants deficient in flavonoid biosynthesis. Plant J 8:659-671.

Smith, T. F., Gaitatzes, C., Saxena, K. and Neer, E. J. (1999) The WD repeat: a common architecture for diverse functions. Trends Biochem. Sci 24: 181–185.

Sommer, H. and Saedler, H. (1986) Structure of the chalcone synthase gene of *Antirrhinum majus*. Mol Gen Genet 202:429-434.

Sommer, H., Carpenter, R., Harrison, B.J. and Saedler, H. (1985) The transposable element *Tam3* of *Antirrhinum majus* generates a novel type of sequence alterations upon excision. Mol Gen Genetics 199(2): 225-231.

Sompornpailin, K., Makita, Y., Yamazaki, M. and Saito, K. (2002) A WD repeat-containing putative regulatory protein in anthocyanin biosynthesis in *Perilla frutescens.* Plant Mol Biol 50: 485–495.

Souer, E., Quattrocchio, F., de Vetten, N., Mol, J. and Koes, R. (1995) A general method to isolate genes tagged by a high copy number transposable element. Plant J 7(4):677–685.

Sparvoli, F., Martin, C., Scienza, A., Gavazzi, G. and Tonelli, C. (1994) Cloning and molecular analysis of structural genes involved in flavonoid and stilbene biosynthesis in grape (*Vitis vinifera* L.). Plant Mol Biol 24: 743–755.

Spelt, C., Quattrocchio, F., Mol, J. and Koes, R. (2002) *ANTHOCYANIN1* of petunia controls pigment synthesis, vacuolar pH, and seed coat development by genetically distinct mechanisms. Plant Cell 14: 2121–2135.

Spelt, C., Quattrocchio, F., Mol, J.N.M. and Koes, R. (2000) *anthocyanin1* of petunia encodes a basic helix–loop–helix protein that directly activates transcription of stuctural anthocyanin genes. Plant Cell 12:1619–1631.

Springob, K., Nakajima, J., Yamazaki, M. and Saito K. (2003) Recent advances in the biosynthesis and accumulation of anthocyanins. Nat. Prod. Rep 20: 288–303.

Stafford, H.A. (1990) Flavonoid Metabolism. CRC Press, Boca Raton, p 298.

Stracke, R., Ishihara, H., Huep, G., Barsch, A., Mehrtens, F., Niehaus, K. and Weisshaar, B. (2007) Differential regulation of closely related R2R3 MYB transcription factors controls flavonols accumulation in different parts of the Arabidopsis thaliana seedling. Plant J 50:660-677.

Stracke, R., Werber, M. and Weisshaar, B. (2001). The R2R3-MYB gene family in *Arabidopsis thaliana*. Curr. Opin. Plant Biol 4: 447–456.

Streisfeld, M.A., Liu, D. and Rausher, M.D. (2011) Predictable patterns of constraint among anthocyanin regulating transcription factors in *Ipomoea*. New Phytol 191(1):264-74.

Sun, Y., Li, H. and Huang, J. (2011) Arabidopsis TT19 Functions as a Carrier to Transport Anthocyanin from the Cytosol to Tonoplasts. Mol Plant 2011:1-14. doi:10.1093/mp/ssr110

Stubbe, H. (1966) Genetik und zytologie von Antirrhinum l. Sect. Antirrhinum. Gustav Fischer, Jena.

Tanaka, Y., Brugliera, F. and Chandler, S. (2009) Recent Progress of flower Colour Modification by Biotechnology. Int J Mol Sci 10: 5350-5369.

Tastard, E., Andalo, C., Giurfa, M., Burrus, M. and Thébaud, C. (2008) Flower colour variation across a hybrid zone in Antirrhinum as perceived by bumblebee pollinators. Arthropod-Plant Interactions 2:237–246.

Tohge, T., Nishiyama, Y., Hirai, M.Y., Yano, M., Nakajima, J., Awazuhara, M., Inoue, E., Takahashi, H., Goodenowe, D.B., Kitayama, M., Noji, M., Yamzaki, M. and Saito, K. (2005) Functional genomic by integrated analysis of metabolome and transcripome of Arabidopsis plants over-expressing an MYB-transcription factor. Plant J 42: 218-235.

Toledo-Ortiz, G., Huq, E. and Quail, P.H. (2003) The Arabidopsis Basic/Helix-Loop-Helix Transcription Factor Family. Plant Cell 15, 1749-1770.

Tornielli, G., Koes, R. and Quattrocchio, F. (2009) The Genetics of Flower Color. Petunia. In Gerats T., and Stromme, J. (2nd ed.) Evolutionary, Developmental and Physiological Genetics. Springer, NY, p 269-299.

Van den Broeck, D., Maes, T., Sauer, M., Zethof, J., De keukeleire, P., D'hauw, M., Van Montagu, M. and Garats, T. (1998) Transposon display identifies individual transpoable element in high copy number liines. Plant J 13:121-129

van der Krol, A.R., Mur, L.A., Beld, M., Mol, J.N.M. and Stuitje A.R.(1990) Flavonoid Genes in Petunia: Addition of a Limited Number of Gene Copies May Lead to a Suppression of Gene Expression. Plant Cell 2: 291-299.

van Nocker, S. and Ludwig, P. (2003) The WD-repeat protein superfamily in Arabidopsis: conservation and divergence in structure and function. BMC Genomics 4: 50.

Van Ooijen, J.W. and Voorrips, R.E. (2001) JoinMap® 3.0, Software for the calculation of genetic linkage maps. Plant Research International, Wageningen, the Netherlands.

van Tunen, A.J., Koes, R.E., Spelt, C.E., van der Krol, A.R., Stuitje, A.R. and Mol, J.N.M. (1988) Cloning of the two chalcone flavanone isomerase genes from *Petunia hybrida*: Coordinate, light-regulated and differential expression of flavonoid genes. EMBO J. 7:1257-1263.

van Tunen, A.J., Mur, L.A., Brouns, G.S., Rienstra, J.D., Koes, R.E. and Mol, J.N.M (1990) Pollen- and Anther-Specific *chí* Promoters from Petunia: Tandem Promoter Regulation of the *chíA* Gene. Plant Cell 2: 393-401.

van Verk, M.C., Bol, J.F. and Linthorst, H.J.M (2011) WRKY Transcription Factors Involved in Activation of SA Biosynthesis Genes. BMC Plant Biology 11:89. http://www.biomedcentral.com/1471-2229/11/89

Vargas, P., Rossello, J. A., Oyama3, R. and Güemes, J. (2004) Molecular evidence for naturalness of genera in the tribe Antirrhineae (Scrophulariaceae) and three independent evolutionary lineages from the New World and the Old. Plant Syst. Evol. 249: 151–172.

Verweij, W., Spelt, C., Sansebastiano, G.D, Vermeer, J., Reale, L., Ferranti, F., Koes, R. and Quattrocchio, F. (2008) An H^+ P-ATPase on the tonoplast determines vacuolar pH and flower colour. Nature Cell Biol 10(12):1456-1462.

Vieira, C.P. and Charlesworth, D. (2002) Molecular variation at the self-incompatibility locus in natural populations of the genera Antirrhinum and Misopates. Heredity 88, 172–181.

Visscher, P.M., Thompson, R., and Haley, C.S. (1996) Confidence intervals in QTL mapping by bootstrapping. Genetics143:1013-1020.

Vogelien, D.L., Hrazdina, G., Reeves, S. and Dougall, D.K. (1990) Phenotypic differences in anthocyanin accumulation among clonally related cultured cells of carrot. Plant Cell, Tissue and Organ Culture 22: 213-222.

Vorobyev, M., Gumbert, A., Kunze, J., Giurfa, M. and Menzel, R. (1997) Flowers through insect eyes. Isr J Plant Sci 45:93–101.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., van de Lee, T., Hornes, M., Frijters, A., Peleman, J. And Kuiper, M. (1995). "AFLP: a new technique for DNA fingerprinting". Nucleic Acids Res. 23 (21): 4407–14.

Waser, N. and Chittka, L. (1998) Bedazzled by flowers. Nature 394: 835-836.

Weil, C.F. and Wessler, S.R. (1990) The Effects of Plant Transposable Element Insertion on Transcription Initiation and RNA Processing. Ann Rev Plant Physiol and Plant Mol Biol 41: 527-552.

Welss, D., van Tunen, A.J., Halevy, A.H., Mol, J.N.M., and Gerats, A.G.M. (1990). Stamen and GA, in the regulation of flavonoid gene expression in the corolla of *Petunia hybrida*. Plant Physiol. 94: 511-515.

Wheldale, M. (1907). The inheritance of flower colour in *Antirrhinum majus*. Proc. R. Soc. Lond. B Biol. Sci. 79, 288–305.

Whibley, A.C. (2004) Molecular and Genetic Variation Underlying the Evolution of Flower Colour in Antirrhinum. PhD thesis, University of East Anglia, UK

Whibley, A.C., Langlade, N.B., Andalo, C., Hanna, A.I., Bangham, A., Thébaud, C. and Coen, E.S. (2006) Evolutionary paths underlying flower color variation in *Antirrhinum*. Science 313, 963-966.

Whitney, H.M., Chittka, L., Bruce, T.J.A. and Glover, B.J. (2009) Conical Epidermal Cells Allow Bees to Grip Flowers and Increase Foraging Efficiency. Curr Biol 19:948-953.

Whitney, H.M. and Glover, B.J (2007) Morphology and development of floral feature recognised by pollinator. Artrophod-Plant interactions 1:147-158.

Whittall, J. B., Voelckel, C. and Hodges, S.A. (2006) Convergence, constraint and the role of gene expression during adaptive radiation: Floral anthocyanins in *Aquilegia*. Mol Eco 15: 4645-4657.

Wilkins, O., Nahal, H., Foong, J., Provart, N. J. and Campbell, M.M. (2009) Expansion and diversification of the *Populus* R2R3-MYB family of transcription factors. Plant Physiol 149: 981–993.

Winkel-Shirley, B., Hanley, S. and Goodman, H.M. (1992) Effects of ionizing radiation on a plant genome: analysis of two *Arabidopsis transparent testa* mutations. Plant Cell 4: 333–347

Xie, D.Y., Jackson, L.A., Cooper, J.D., Ferreira, D., and Paiva, N.L. (2004) Molecular and biochemical analysis of two cDNA clones encoding dihydoflavonol-4-reductase from *Medicago truncatula*. Plant Physiol 134, 979–994.

Yamazaki, M., Yamagishi, E., Gong, Z., Fukuchi-Mizutani, M., Fukui, Y., Tanaka, Y., Kusumi, T., Yamaguchi, M. and Saito, K. (2002) Two flavonoid glucosyltransferase from *Petunia hybrida*: molecular cloning, biochemical properties and developmentally regulated expression. Plant Mol Biol 48(4):401-411.

Yang, C.C. (2007) Identifying Genes that Underlie Evolution of *Antirrhinum* species. PhD thesis, University of Edinburgh.

Yazaki, K. (2005) Transporters of secondary metabolites. Curr Opin Plant Biol 8: 301-307.

Yonekura-Sakakibara, K., Nakayama, T., Yamzaki, M. and Saito, K. (2009) Modification and Stabilization of Anthocyanins, in: Anthocyanins, (eds. Gould, K. et al.), Springer, NY, p 69-190.

Zhao, J., and Dixon, R. A. (2010) The 'ins' and 'outs' of flavonoid transport. Trends in Plant Science, 15 (2): 72-80. DOI:10.1016/j.tplants.2009.11.006.

Zhao, J. and Dixon, R.A. (2009) MATE transporters facilitate vacuolar uptake of epicatechin 3′-O-glucoside for proanthocyanidin biosynthesis in *Medicago truncatula* and *Arabidopsis*. Plant Cell 21: 2323–2340.

Zhao, J., Huhman, D., Shadle, G., He, X.Z., Sumner, L.W., Tang, Y. and Dixon, R.A. (2011) *MATE2* Mediates Vacuolar Sequestration of Flavonoid Glycosides and Glycoside Malonates in *Medicago truncatula*. Plant Cell 23: 1536–1555.

Zhao, M., Morohashi, K., Hatlestad, G., Grotewold, E. and Lloyd, A. (2008) The TTG1–bHLH–MYB complex controls trichome cell fate and patterning through direct targeting of regulatory loci. Development 135: 1991–1999.

Zufall, R.A, and Rausher, M.D (2004) Genetic changes associated with floral adaptation restrict future evolutionary potential. Nature 428: 847-850.